U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746-4304

Office Note 435

PROPOSED SEMI-IMPLICIT ADAPTATIONS OF TWO LOW-STORAGE RUNGE-KUTTA SCHEMES. PART I: THEORETICAL FORMULATION AND STABILITY ANALYSIS

R. James Purser*
Science Applications International Corp., Beltsville, Maryland

August 2001

* email: jim.purser@noaa.gov

Abstract

This note describes two novel semi-implicit schemes, intended primarily for application to forward-trajectory semi-Lagrangian models. The explicit components of the algorithms are significantly more accurate than the commonly used leapfrog method and are based on low-storage versions of the third- and fourth-order Runge-Kutta technique. The fast modes only are adjusted at each of the stages of the Runge-Kutta cycle using a slightly dissipative method that generalizes the trapezoidal implicit scheme either by first-order accurate de-centering involving the two time levels straddling the active time interval, or by a second-order accurate extension involving also one earlier time level. These semi-implicit extensions should be reasonably straight-forward to apply and, with appropriate choices of the schemes' de-centering parameters, will lead to robustly stable methods applicable to both hydrostatic and nonhydrostatic atmospheric models. In this part of the study we review some of the properties of the methods in the idealized dynamical system consisting of a simple damped linear oscillator. This system allows us to examine the robustness of different de-centering strategies to errors in the specification of the modal frequencies. From these experiments it becomes apparent that a restriction to de-centering strategies that preserve formal second-order temporal accuracy can never provide a sufficient degree of robustness for those implicit modes that have low frequencies. Although, for such modes, the unmodified explicit Runge-Kutta schemes would remain stable, we have found no satisfactory way of relaxing or 'diluting' the implicit adjustments for these slower modes that is capable of avoiding the inherent lack of robustness while preserving second-order accuracy. Fortunately, the simpler first-order decentering strategy *does* recover robustness throughout the spectrum of implicit modes and retains this property when a dilution of the implicit adjustments is carried out for the slowest of these modes so that the time integration of them can be made to revert progressively towards the more accurate Runge-Kutta schemes. A sequel to this article will describe the practical aspects of applying the proposed techniques to actual nonlinear numerical models, where we can expect the semi-Lagrangian versions, unrestricted by the usual advective CFL limit, to out-perform the leapfrog method in terms of accuracy while offering opportunities for reducing the storage requirements and computational costs.

1. Introduction

In a semi-implicit Eulerian numerical model it is very difficult to find a basic time integration scheme, for the explicit dynamical components, that is superior to the traditional time-filtered leapfrog method. This is because the time step is dictated essentially by the advective Courant-Friedrichs-Lewy (CFL) criterion of the particular integration scheme, and the leapfrog scheme's CFL restriction is the most lenient of all the well-known integration methods. In a semi-Lagrangian model, other factors determine how large a time step should be. For example, a compromise is usually made between the insertion rate of physical or topographic forcing sufficiently high to capture the physical processes accurately and the longer time step permitted for a semi-Lagrangian model purely on the basis of numerical stability. Since the time steps of a semi-Lagrangian model are normally significantly longer than those of a comparable Eulerian

model, it may be valid to pay greater attention to the minimization of time-trunction errors in the semi-Lagrangian case, but the advective CFL condition certainly has no bearing on the time step choice. In developing semi-Lagrangian models it is therefore reasonable to question whether there are faster, or more accurate time stepping methods than the semi-implicit time-filtered leapfrog method. This article will attempt to answer this question affirmatively and provide two suitable schemes, based on semi-implicit adaptations of low-storage Runge-Kutta methods.

We will begin the next section by reviewing relevant features of modern semi-Lagrangian frameworks and some of the available families of time-integration methods, together with what (if anything) these potentially applicable method have to offer in the context of the semi-Lagrangian models. The time-filtered leapfrog method will be taken as the effective benchmark scheme with regard to both accuracy and storage economy since, among the standard explicit integration methods, it possesses the largest range of stability in the frequency domain for the purely oscillatory modes that are of paramount importance in most geophysical flows. It is essential that a robust semi-implicit form of any scheme should be demonstrated before that scheme be deemed acceptable. Often one finds that there is a trade-off one can make within the same family of schemes between storage economy and formal accuracy. In fact, having narrowed down our recommended field of contenders, we find that this trade-off applies between the two choices we finally offer: one is a third-order method actually requiring less storage than the leapfrog, whose explicit form was proposed by Williamson (1980); the other method, whose explicit form was proposed by Gill (1951), achieves fully fourth-order accuracy (for the explicitly integrated components) but requires as much storage as the leapfrog. When memory is available, it is the fourth-order method that we recommend, since the automatic occurrence of a division of the principal time step into just two equal sub-intervals leads to some valuable simplifications in practice.

Section 2 reviews the characteristics of some of the available time integration schemes to provide the motivation for us to consider the low-storage Runge-Kutta methods. Section 3 describes the explicit methods of Williamson (1980) and Gill (1951), while the proposed semi-implicit extensions of these two methods are defined and evaluated in Section 4. The concluding section of this part of the study discusses the implication of these experiments for the choice of scheme in a full nonlinear primitive equations model, whose detailed examination will be the topic of Part II of this investigation.

2. Integration methods in the semi-Lagrangian framework

The traditional approach to numerical weather prediction, employing the Eulerian equations, is gradually being superseded in most prediction centers by the potentially more efficient semi-Lagrangian methods. Reintroduced in the early nineteen eighties by Robert (1981, 1982) in combination with the semi-implicit method (e.g., Robert et al. 1972), and also adapted by Bates and McDonald (1982) for split-explicit treatment of fast modes, the semi-Lagrangian method, as mentioned earlier, offers the opportunity to achieve stable numerical integration using timesteps unrestricted by the ordinary advective Courant-Friedrichs-Lewy (CFL) stability criterion. It has been shown by Purser and Leslie (1988) that there are benefits to be gained by adopting high-order spatial numerics in a semi-Lagrangian model. The so-called

'cascade' method of grid-to-grid interpolation was proposed by Purser and Leslie (1991) as a way to enhance the efficiency of the semi-Lagrangian interpolations needed at each time step when high-order numerics are employed in the construction of the interpolation operators. The cascade method reduces the three-dimensional interpolation problem to a sequence of three separate sets of one-dimensional grid-to-grid interpolations through a sequence of grids constructed as hybrids between the Eulerian and Lagrangian coordinate frameworks.

One consequence of adopting the cascade method is that interpolation from forward trajectories becomes virtually no more complicated than the more conventional interpolation from Eulerian to Lagrangian coordinates used in most existing semi-Lagrangian models that employ backward trajectories. The forward trajectory approach has the notable advantage of lending itself naturally to the application of a much wider variety of time integration schemes, avoiding the iterative adjustments of trajectory origins that so often characterize the backward trajectories implementation of some time discretization methods. In particular, integration schemes possessing a formal order of accuracy greater than the commonly adopted first- or second-order methods can be considered for use within a forward-trajectory semi-Lagrangian model. This is a way to guard against the possibility of significant time truncation errors that might occur as a consequence of being able to run the model with time steps much larger than those typical of a comparable Eulerian model. By using forward trajectories, essentially all the standard numerical time integration methods for ordinary differential equations (ODEs) are immediately made available for application to the Lagrangian-frame dynamics of our numerical weather prediction models.

There is at least some evidence from the studies of Purser and Leslie (1994, 1996) that small, but nevertheless significant, benefits to forecast accuracy (as measured by root-mean-square fits to a high resolution control forecast) result from the use of high-order integration techniques in a semi-Lagrangian model. Durran (1999) has nicely summarized the important distinguishing properties of several standard time integration methods (see his Tables 2.1 and 2.2, pp.68–69). It is normally impractical to use the leapfrog benchmark method in its pure form, owing to the presence of an undamped computational mode for oscillatory dynamics which becomes unstable in the presence of the smallest amount of dissipation. It is usual to damp the computational mode by slightly modifying the pure second-order leapfrog by the inclusion of a time filter (Robert 1966, Asselin 1972). Unfortunately, this formally reduces the order of accuracy (and slightly reduces the range of stability). However, with the typically small size of the adjustable smoothing parameter used in practice (the NCEP global model currently uses $\gamma = 0.04$ in the notation of Haltiner and Williams 1980, or $\nu = 0.08$ in the notation of Asselin 1972), the coefficient of first-order error is very small; in most respects, the modified scheme therefore still behaves like a fully second-order method. The filtered leapfrog requires a pair of auxiliary storage fields (making three fields in total when counting the dynamical state variables themselves), which is considered a reasonable storage cost for an operational model. As shown by Robert et al. (1972), the leapfrog method can be converted into a semi-implicit scheme without great difficulty, assuring the method's continued utility for both Eulerian and semi-Lagrangian numerical weather prediction models.

A perusal of Durran's tables of standard schemes suggests that few, if any, of the other methods of second-order accuracy qualify for service in weather prediction models; only the Magazenkov (1980) scheme, which alternates a leapfrog step with the Adams-Bashforth second-

order method, and the Kurihara (1965) explicit leapfrog-trapezoidal schemes, are listed as being stable for oscillatory modes. Both of these schemes have inferior stability ranges compared to the leapfrog scheme once time step sizes are adjusted to equalize the computational costs associated with the 'function evaluations' of state tendencies. Both schemes have storage requirements identical to the those of the filtered-leapfrog benchmark. It does not seem that the slightly improved accuracy of these schemes is sufficient to justify the effort and complexity involved in forming semi-implicit forms of them that could potentially replace the semi-implicit time-filtered leapfrog.

Turning to higher orders of accuracy, one finds that many of the classical numerical integration schemes require several additional auxiliary fields of storage. Clearly, such schemes are much less attractive, especially when the additional fields incur not simply an extra storage burden, but, by having to be saved and interpolated from one time level to the next in a semi-Lagrangian model, also incur a significant additional computational burden. The various families of high-order multistep methods, to which belong the Adams-Bashforth methods (Gear 1971), fall into this unfortunate category. For example, the third-order Adams-Bashforth method has excellent accuracy and moderately good stability characteristics and has been recommended for meteorological models largely for these reasons by Durran (1991). A semi-implicit adaptation of this method was implemented in a semi-Lagrangian model by Purser and Leslie (1994) with good results, judged by objective criteria and by the subjective interpreation of numerous forecasts. However, the storage required by the third-order Adams-Bashforth method totals four fields of data, which exceeds the requirements of the leapfrog benchmark. Similarly, there are extensions to the Adams-Bashforth family that apply to the second-derivative equations for Lagrangian trajectories, combining both the kinematic and momentum equations. These schemes become the Störmer (1907) schemes in the absence of a Coriolis parameter, but the more general forms are referred to as 'Generalized Adams-Bashforth' schemes by Purser and Leslie (1996). These schemes also require several auxiliary storage fields if a high-order of accuracy is to be attained.

Another family of classical methods for integrating ODEs to higher orders of accuracy are the 'one-step', including Runge-Kutta methods (Gear 1971). In general, these methods also require a number of auxiliary storage fields, although the intermediate fields of data are not passed on from one cycle of time steps to the next. Attempts have been made to address the problem of storage economy. Lorenz (1971) proposed a family of Runge-Kutta-like 'N-cycle' schemes in which one is required to provide only a single auxiliary storage field to hold successive linear combinations of present and previous tendencies of the state. The N individual time steps that make up each cycle are of equal size in his schemes. For each $N > 2$, there exist two distinct N-cycle schemes that achieve a formal accuracy of Nth-order when applied to problems in which the tendencies are exactly linear functions of the state variables, whereupon, the methods $N \leq 4$ become equivalent to the corresponding Runge-Kutta schemes. These schemes are defined in Appendix A.

It is instructive to compare the stability regimes of a few of the methods we have discussed in the context of the simple differential equation,

$$\frac{d\psi}{dt} = J\psi, \tag{2.1}$$

which represents an oscillator when J is purely imaginary, and represents frictional damping or diffusion when J is real and negative. A meteorological example whose description might involve an equation of this kind is the dynamical behavior of inertial oscillation in the nocturnal boundary layer (S. G. Gopalakrishnan, personal communication). Figure 1 shows the regions of the complex space of $J\delta t$ (bounded by the closed curves) where the filtered-leapfrog, the third-order Adams-Bashforth scheme and the 4th-order Runge-Kutta (or Lorenz 4-cycle) schemes remain stable, the δt being the average time between function evaluations rather than a true time step, in the case of the Runge-Kutta method. Appendix B describes how these curves are obtained in general (and reminds the reader of the defining formula for the third-order Adams-Bashforth method examined in Fig. 1). A reduction in the stable frequency range of these higher order methods compared to the leapfrog is evidently compensated to some degree by their better stability in dissipative situations. Figure 2 shows the stability regions achieved for all of the N-cycle methods for the range $N \in [1, 5]$. Up to fourth-order, these are equivalent to classical Runge-Kutta stability regions. Although it is not always visually obvious unless the scale of the real component is expanded, the stability curves of $N = 1$, 2 and 5 do *not* enclose a finite stretch of the imaginary axis near the origin and are consequently *unstable* for even the slowest oscillatory modes. When we consider the true Runge-Kutta methods below, we shall therefore confine our consideration to third- and fourth-order schemes. An attempt to generalize Lorenz's maximally storage-frugal method to the second-derivative momentum/kinematic equations of a semi-Lagrangian model was given by Purser and Leslie (1997). However, all of the supposedly higher-order N-cycle schemes are *not* fully Nth-order accurate for generic tendencies that are *nonlinear* functions of the state variables. If we apply the two alternative forms of the Lorenz 4-cycle (see Appendix A for their definitions) to the dynamics implied by a system of central force law equations in which the force is *not* a linear function of radial distance, we obtain truncation errors of phase that can be quite large compared to those obtained with the corresponding classical Runge-Kutta scheme (also defined in Appendix A), even when the true 'orbits' are perfectly circular and the magnitude of the true force is remaining constant. The celebrated case of the inverse-square force law, although of no specifically meteorolgical interest, was coded as a convenient test example because the general solutions (Kepler orbits) remain analytically simple even when they are not circular. In this study, however, we only show the attempts at simulating the circular orbit solutions with the above-mentioned three methods. Figure 3 compares these methods together with the 'true' solution obtained by integrating with very short time steps. We observe that the phase errors of the two alternative N-cycle schemes are approximately equal and opposite. Lorenz suggested alternating the two N-cycles in order to achieve the full Nth-order of accuracy over groups of cycles, but the efficacy of this strategy in a highly nonlinear numerical weather prediction model is extremely doubtful.

Lorenz imposes two restrictions: first, that time steps of each stage of the cycle are equal; second, that only a single auxiliary storage field is provided for the intermediate computations. Relaxing these restrictions enables fully Nth-order schemes, the proper Runge-Kutta family, to be considered. Wicker and Skamarock (2001) have recently examined a third-order Runge-Kutta scheme whose time step subdivisions are convenient for application to the slow modes of the split-explicit Eulerian dynamical core of the fully compressible nonhydrostatic Weather Research and Forecasting (WRF) model (see also Skamarock et al. 2001) and they report encouraging results. We shall not consider methods of orders greater than four; even a glance

at Fig. 2 strongly suggests that we should not expect a higher order explicit Runge-Kutta scheme to produce an effective stability range for oscillatory modes better than that produced by the fourth-order method and it can be shown rigorously (Butcher, 1965) that truly fifth-order Runge-Kutta schemes require no less than six stages. Williamson (1980), extending earlier work described by Kopal (1961), examined the problem of implementing low-storage Runge-Kutta methods. From a continuously parameterized family of possible schemes requiring a total of only two storage fields, one particular method, henceforth referred to here as the Williamson method, was recommended. Note that the storage economy of such a method improves upon that of the leapfrog scheme. Another, and perhaps better known low-storage Runge-Kutta method is the fourth-order scheme devised by Gill (1951), which requires a total of three storage fields. Because the minimization of storage requirements is always of critical concern in a numerical weather simulation or prediction model, these members of the Runge-Kutta family are of particular interest to us. We shall therefore devote the next sections to discussions of these two explicit schemes, their application to semi-Lagrangian models and the further adaptations to them that convert them to semi-implicit methods.

3. The Runge-Kutta methods of Williamson and Gill

The three-stage algorithm of Williamson (1980) and the four-stage algorithm of Gill (1951) belong to the class of Runge-Kutta (RK) schemes (e.g., see Butcher 1987). The nice thing about these schemes is that they achieve third- and fourth-order accuracy in time with only quite modest memory requirements. In the case of the third-order Williamson scheme, only two storage fields are needed while, in the case of Gill's scheme, three fields are needed. In each case, one of these storage fields is reserved to hold the state variables themselves. The generic form of an explicit Runge-Kutta method is explained in Appendix A. Table 1 summarizes the Williamson algorithm adapted to the context of a semi-Lagrangian model capable of handling the nonstandard case of forward trajectories. Provision is made in this summary for the inclusion of adjustment increments (denoted ψ_{adj}) which would be needed in a split-semi-implicit modification of the scheme, although we defer discussion about the creation of suitable forms for these terms until later; the explicit Williamson schemes result from setting all the terms ψ_{adj} to zero. Williamson particularly recommends the scheme defined in Table 1 with the coefficients:

$$R_0 = \frac{1}{3}, \tag{3.1}$$

$$R_1 = \frac{15}{16}, \tag{3.2}$$

$$R_2 = \frac{8}{15}, \tag{3.3}$$

$$Q_1 = -\frac{25}{16}, \tag{3.4}$$

$$Q_2 = -\frac{17}{25}. \tag{3.5}$$

Other schemes of the same kind can be found by the methods described in Appendix A.

Table 2 summarizes Gill's fourth-order algorithm, similarly modified to accommodate semi-implicit terms. In this case, an additional storage field must now be provided for holding some of the intermediate results. The two numerical constants, A and B, we use to define this version of Gill's scheme are:

$$A = 2 - \sqrt{2} \approx 0.5857864376269049512 \quad ,$$
$$B = 1 + \sqrt{2} \approx 2.4142135623730950488 \quad .$$

TABLE 1. SEMI-LAGRANGIAN FORM OF WILLIAMSON'S RK3 ALGORITHM

To/from Archive	Stage	Operation	Field-1	Field-2
$\psi_t \rightarrow \psi^0$	1	$E^0 = R_0 \mathcal{F}(\psi^0)\delta t$	ψ^0	E^0
		$\psi^1_{[0]} = \psi^0 + E^0 + \psi^1_{\text{adj}[0]}$	$\psi^1_{[0]}$	
		$(\psi^1, E^0_{[1]}) = \mathcal{I}_{[1\leftarrow 0]}(\psi^1_{[0]}, E^0)$	ψ^1	$E^0_{[1]}$
	2	$E^1 = R_1 \mathcal{F}(\psi^1)\delta t + Q_1 E^0_{[1]}$		E^1
		$\psi^2_{[1]} = \psi^1 + E^1 + \psi^2_{\text{adj}[1]}$	$\psi^2_{[1]}$	
		$(\psi^2, E^1_{[2]}) = \mathcal{I}_{[2\leftarrow 1]}(\psi^2_{[1]}, E^1)$	ψ^2	$E^1_{[2]}$
	3	$E^2 = R_2 \mathcal{F}(\psi^2)\delta t + Q_2 E^1_{[2]}$		E^2
		$\psi^3_{[2]} = \psi^2 + E^2 + \psi^3_{\text{adj}[2]}$	$\psi^3_{[2]}$	$\cdots$
$\psi_{t+\delta t} \leftarrow \psi^3$		$\psi^3 = \mathcal{I}_{[3\leftarrow 2]}(\psi^3_{[2]})$	ψ^3	$\cdots$

In tables 1 and 2, the state variables are represented collectively by ψ, with ψ^a denoting their values on the Eulerian standard grid at stage a of the time step cycle. For variables identified with stage a, but held on the Lagrangian trajectories that intersect the Eulerian standard grid only at another stage, b, of the cycle, the notation '$\psi^a_{[b]}$' is used. The computed tendencies of ψ at stage a are denoted by F^a and are always first estimated by the application of the nonlinear 'tendency operator' $\mathcal{F}$ to state ψ^a. The algorithms tabulated above assume the existence of a procedure which can perform the interpolations between the Eulerian and Lagrangian grids in the nonstandard direction — that is, *from Lagrangian to Eulerian* grids. This interpolation operator, which is normally linear in the variables being interpolated, but can also be 'shape-preserving', is here represented symbolically by $\mathcal{I}_{[b\leftarrow a]}$, where a denotes the stage at which the Lagrangian source grid previously coincided with the standard Eulerian grid and b denotes the stage of the time stepping cycle labelling the target Eulerian standard grid. In stages 2 and 4 of the Gill algorithm set out in Table 2 the time level is a repeat of the preceding time level, so the interpolation operator involves only a very small set of relative displacements. This enables the interpolation operator at these stages to be represented by a formal linearization about zero displacement, a special instance we signify in this table by the notation, $\mathcal{I}^*_{[b\leftarrow a]}$. In effect, the advection at these stages reverts to an Eulerian treatment. The time stepping cycle itself occupies a period of time δt beginning at stage 0.

TABLE 2. SEMI-LAGRANGIAN FORM OF GILL'S RK4 ALGORITHM

To/from Archive	Stage	Operation	Field-1	Field-2	Field-3
$\psi_t \rightarrow \psi^0$	1	$\frac{1}{2}F^0 = \frac{1}{2}\mathcal{F}(\psi^0)\delta t$	ψ^0	...	$\frac{1}{2}F^0$
		$G^1_{[0]} \equiv E^0 = \frac{1}{2}F^0$		$G^1_{[0]} \equiv E^0$	...
		$\psi^1_{[0]} = \psi^0 + E^0 + \psi^1_{\mathrm{adj}[0]}$	$\psi^1_{[0]}$		...
		$(\psi^1, G^1) = \mathcal{I}_{[1\leftarrow 0]}(\psi^1_{[0]}, G^1_{[0]})$	ψ^1	G^1	...
	2	$\frac{1}{2}F^1 = \frac{1}{2}\mathcal{F}(\psi^1)\delta t$			$\frac{1}{2}F^1$
		$E^1 = A(\frac{1}{2}F^1 - G^1)$		E^1	
		$\psi^2_{[1]} = \psi^1 + E^1 + \psi^2_{\mathrm{adj}[1]}$	$\psi^2_{[1]}$		
		$G^2_{[1]} = \frac{1}{2}F^1 - \frac{1}{2}AE^1$		$G^2_{[1]}$	...
$\psi_{t+\delta t/2} \leftarrow \psi^2$		$(\psi^2, G^2) = \mathcal{I}^*_{[2\leftarrow 1]}(\psi^2_{[1]}, G^2_{[1]})$	ψ^2	G^2	...
	3	$\frac{1}{2}F^2 = \frac{1}{2}\mathcal{F}(\psi^2)\delta t$			$\frac{1}{2}F^2$
		$E^2 = \frac{1}{2}F^2 + B(\frac{1}{2}F^2 - G^2)$		E^2	
		$\psi^3_{[2]} = \psi^2 + E^2 + \psi^3_{\mathrm{adj}[2]}$	$\psi^3_{[2]}$		
		$G^3_{[2]} = \frac{1}{2}F^2 + B(E^2 - \frac{1}{2}F^2)$		$G^3_{[2]}$	...
		$(\psi^3, G^3) = \mathcal{I}_{[3\leftarrow 2]}(\psi^3_{[2]}, G^3_{[2]})$	ψ^3	G^3	...
	4	$\frac{1}{2}F^3 = \frac{1}{2}\mathcal{F}(\psi^3)\delta t$			$\frac{1}{2}F^3$
		$E^3 = \frac{1}{3}(\frac{1}{2}F^3 - G^3)$		E^3	...
		$\psi^4_{[3]} = \psi^3 + E^3 + \psi^4_{\mathrm{adj}[3]}$	$\psi^4_{[3]}$	...	...
$\psi_{t+\delta t} \leftarrow \psi^4$		$\psi^4 = \mathcal{I}^*_{[4\leftarrow 3]}(\psi^4_{[3]})$	ψ^4	...	...

In Williamson's algorithm, stages 1 and 2 update the intermediate state to times $t + c_1\delta t$ and $t + c_2\delta t$, where, for the particular scheme recommended,

$$c_1 = \frac{1}{3}, \tag{3.6}$$

$$c_2 = \frac{3}{4}. \tag{3.7}$$

The final stage brings the solution up-to-date at $t + \delta t$.

In the case of Gill's fourth-order algorithm, stages 1 and 2 are both identified with the middle of the period, or $\delta t/2$ from the beginning, while stages 3 and 4 are identified with the end of the cycle, or δt from the beginning. In the notation by which Runge-Kutta schemes are defined in Appendix A, this scheme therefore has,

$$c_1 = \frac{1}{2}, \tag{3.8}$$

$$c_2 = \frac{1}{2}, \tag{3.9}$$

$$c_3 = 1. \tag{3.10}$$

The final state of the cycle, ψ^4, then becomes the initial state, ψ^0, of the following cycle.

Tables 1 and 2 make provision for the possibility of semi-implicit adjustments at each stage. The following section offers suggestions about how this might be done.

4. Semi-implicit treatment

For oscillatory or dissipative systems, integration for time step, $c_1\delta t$, with $F_t = \mathcal{F}(\psi(t))\delta t$ by the implicit scheme:

$$\psi(t + c_1\delta t) = \psi(t) + W_0 F(t) + W_1 F(t + c_1\delta t), \tag{4.1}$$

is normally stable when $W_1 \geq W_o$ and $W_0 + W_1 = c_1$, and achieves second-order accuracy when $W_0 = W_1 = c_1/2$ (the 'trapezoidal method'). If we write,

$$(W_0, W_1) = \frac{c_1}{2}\,(1 - a, 1 + a)\,, \tag{4.2}$$

where a is the first-order de-centering parameter, the second-order trapezoidal method corresponds to $a = 0$, the Euler-backward method corresponds to $a = 1$, and other first-order implicit methods correspond to intermediate choices of a. When three times levels, $(t, t + c_1\delta t, t + c_2\delta t)$, are available, it is sometimes useful to generalize the implicit scheme so that the fast oscillatory modes can be positively damped while remaining asymptotically accurate to second order. This is accomplished by introducing an additional decentering parameter, b. Then, in the more general scheme,

$$\psi(t + c_2\delta t) = \psi(t + c_1\delta t) + W_0 F(t) + W_1 F(t + c_1\delta t) + W_2 F(t + c_2\delta t), \tag{4.3}$$

where,

$$(W_0, W_1, W_2) = \frac{c_2 - c_1}{2}\left(b - b\frac{c_1}{c_2}, 1 - a - b, 1 + a + b\frac{c_1}{c_2}\right), \tag{4.4}$$

a definite damping of the oscillatory modes occurs even when the first-order de-centering parameter a vanishes, provided the second-order de-centering parameter b is positive.

We propose to adopt such schemes for the implicitly treated components of a numerical integration whose explicit modes are dealt with using either the Williamson or the Gill algorithms described in the previous section. The de-centering parameters ensure a positive degree of dissipation even for those modes which should really be neutrally stable; in this way, we confer upon our schemes an inherent robustness that the trapezoidal scheme alone would not possess. However, the second-order decentering (involving three time levels) is only convenient to apply during the middle subdivision of Williamson's algorithm or during the second half of Gill's algorithm, since it is only at these stages that we have convenient access to just the three time levels that are needed, while still being able to extract from the auxiliary storage fields the necessary tendency terms. A greater degree of robustness is obtained, at the expense of incurring a first-order temporal error in the fast modes, by using a small positive de-centering

parameter, a, in any or all of the stages of the modified Williamson or Gill schemes. Since the de-centering parameters may differ in each of the stages, we shall employ the index of each stage as a distinguishing subscript. The implicit tendencies must be estimated on the basis of a suitable linearization about the latest explicitly available state; for example, in stage 1 of either the Williamson or Gill algorithms, we might employ the approximation (dropping the trajectory subscripts for clarity):

$$\mathcal{F}(\psi^1 + \psi^1_{\rm adj})\delta t \approx F^0 + J^*(\psi^1 + \psi^1_{\rm adj} - \psi^0), \tag{4.5}$$

where J^* denotes an approximation to the Jacobian operator J encoding the sensitivity, for the fast modes, of $F \equiv \mathcal{F}(\psi)\delta t$ with respect to changes in ψ:

$$J^* \approx J, \tag{4.6}$$

where

$$dF = Jd\psi. \tag{4.7}$$

We now introduce the proposed modifications to the Williamson and Gill algorithms that make them semi-implicit, and investigate their properties in highly idealized dynamical systems designed to reveal some aspects of these schemes' robustness to mis-specification of modal frequencies.

(a) Semi-implicit modified Williamson scheme

We use the two-level semi-implicit treatment for stages 1 and 3 but the three-level form (i.e., with a b-parameter) for the middle stage. In principle, this should allow a fully second-order implicit scheme (vanishing a-parameters) with a definite dampening of the fast modes, and therefore, a built-in robustness to some extent. The adjustments at each of the three stages are given by:

$$E^0 + \psi^1_{\rm adj} = \frac{1-a_1}{6}F^0 + \frac{1+a_1}{6}\left[F^0 + J^*(E^0 + \psi^1_{\rm adj})\right], \tag{4.8}$$

$$E^1 + \psi^2_{\rm adj} = W_E E^1 + W_F F^1 + \frac{5}{24}(1 + a_2 + \frac{4b}{9})\left[F^1 + J^*(E^1 + \psi^2_{\rm adj})\right], \tag{4.9}$$

$$E^2 + \psi^3_{\rm adj} = \frac{1-a_3}{8}F^2 + \frac{1+a_3}{8}\left[F^2 + J^*(E^2 + \psi^3_{\rm adj})\right], \tag{4.10}$$

where, since from Table 1 we have,

$$E^1 = \frac{(-25F^0 + 45F^1)}{48}, \tag{4.11}$$

the weights W_E and W_F become consistent with (4.4) if they are defined:

$$W_E = -\frac{2b}{9}, \tag{4.12}$$

$$W_F = \frac{5(1-a_2)}{24}. \tag{4.13}$$

By manipulating these equations we can place only the ψ_{adj} terms on the left-hand sides:

$$\psi^1_{\text{adj}} = (1 - W^1_1 J^*)^{-1} W^1_0 F^0 - E^0, \tag{4.14}$$

$$\psi^2_{\text{adj}} = (1 - W^2_2 J^*)^{-1} (W_E E^1 + W^2_1 F^1) - E^1, \tag{4.15}$$

$$\psi^3_{\text{adj}} = (1 - W^3_3 J^*)^{-1} W^3_2 F^2 - E^2, \tag{4.16}$$

where the new set of weights are:

$$W^1_0 = \frac{1}{3}, \tag{4.17}$$

$$W^1_1 = \frac{1 + a_1}{6}, \tag{4.18}$$

$$W_E = -\frac{2b}{9}, \tag{4.19}$$

$$W^2_1 = \frac{5}{12} + \frac{5b}{54}, \tag{4.20}$$

$$W^2_2 = \frac{5}{24}\left(1 + a_2 + \frac{4b}{9}\right), \tag{4.21}$$

$$W^3_2 = \frac{1}{4}, \tag{4.22}$$

$$W^3_3 = \frac{1 + a_3}{8}. \tag{4.23}$$

In the context of a complete model, the Jacobian operator J^* becomes a matrix of spatial partial derivative operators structured so that, with further manipulations, each adjustment equation involves the solution of a Helmholtz equation. It is instructive to examine the stability regions of schemes of this kind when applied to one of the simplest of dynamical systems, namely, the linear damped oscillator equation, (2.1), for the case of a unit time step, $\delta t = 1$. In this case, the Jacobian J reduces simply to the complex frequency. We have at our disposal the four de-centering parameters, a_1, a_2, a_3 and b. For each value of the assumed frequency J^* we may then plot the boundary of the region of complex J (the *true* complex frequency) for which this particular semi-implicit modification of the Williamson method remains stable. Fig. 4 shows the neutral stability curve in the complex J plane of an unmodified explicit third-order Runge-Kutta scheme (such as Williamson's scheme) for reference (the dotted curve that is symmetric about the real-axis) and the corresponding curves for a semi-implicit modification of Williamson's scheme where all the a's and b vanish. The three representative values of the assumed frequency are: $J^* = i, 3i, 5i$, where $i = \sqrt{-1}$. Note that this scheme exhibits essentially no robustness to erroneous specification of the assumed frequencies, J^*, since the curves of stability intersect these assumed frequencies, bringing unstable regions of the imaginary frequency axis arbitrarily close to the assumed values in each case. Fig. 5 shows the improved robustness obtained by keeping the a coefficients zero but setting $b = .5$. However, the low-frequency implicit modes (exemplified here by the curve for $J^* = i$) still possess little margin for error, and the situation is not much improved for these modes when we further double the parameter b (Fig. 6). When we do not employ the second-order de-centering parameter (that is, we reset $b = 0$) but set all

the a parameters to $a = 0.5$, then, as shown in Fig. 7, all the implicit modes become very robust to mis-specification of their frequencies. In cases where we would be more comfortable having a smaller coefficient of first-order time truncation even for these supposedly unimportant implicit modes, the scheme with $a_1 = a_2 = a_3 = .2$ and $b = .5$, shown in Fig. 8 might be preferable, although the robustness of the slower implicit modes is noticeably reduced in comparison to the scheme illustrated in Fig. 7 and, arguably, the value of having a non-vanishing b in this case is marginal at best.

(*b*) *Semi-implicit modified Gill scheme*

In a manner resembling our treatment of Williamson's third-order Runge-Kutta scheme, we may modify the Gill fourth-order Runge-Kutta algorithm to achieve numerical stability and robustness for modes too fast to be handled by the explicit dynamics. However, in this case, the successive stages do not all involve a positive progression of time; the result of stage 2 is contemporary with that of stage 1, while the result of final stage 4 is contemporary with the result of stage 3. The approach we adopt is to apply the two-level implicit scheme at stage 1; attempt to keep the fast components unchanged at stage 2; apply the three-level implicit scheme (should a positive b parameter be adopted) at stage 3; keep the fast components unchanged again at stage 4. The three-level weighting of (4.4) for the case of equal intervals of $\delta t/2$ takes the form,

$$(W_0, W_1, W_2) = \left(\frac{b}{8}, \frac{1-a-b}{4}, \frac{1+a}{4} + \frac{b}{8}\right). \tag{4.24}$$

Algebraically, the proposed modification of Gill's method therefore translates to:

$$E^0 + \psi^1_{\rm adj} = \frac{1-a_1}{4} F^0 + \frac{1+a_1}{4}\left[F^0 + J^*(E^0 + \psi^1_{\rm adj})\right], \tag{4.25}$$

$$E^1 + \psi^2_{\rm adj} = 0, \tag{4.26}$$

$$E^2 + \psi^3_{\rm adj} = W_E E^2 + W_F F^2 + \frac{(1 + a_3 + \frac{1}{2}b)}{4}\left[F^2 + J^*(E^2 + \psi^3_{\rm adj})\right], \tag{4.27}$$

$$E^3 + \psi^4_{\rm adj} = 0. \tag{4.28}$$

In this case, we deduce from Table 2 that

$$E^2 = \left(\frac{1}{2} - \frac{\sqrt{2}}{2}\right) F^0 - F^1 + \left(1 + \frac{\sqrt{2}}{2}\right) F^2. \tag{4.29}$$

In order to conform to the weighting, (4.24), we first note that the terms F^1 and F^2 of (4.29) are two approximations to the same central-time forcing, associated with the weight, W_1, of (4.24). The consistent definitions of W_E and W_F of (4.27) are then:

$$W_E = -\frac{(1+\sqrt{2})b}{4}, \tag{4.30}$$

$$W_F = \frac{1-a_3}{4} + \frac{\sqrt{2}b}{8}. \tag{4.31}$$

As in the case of the semi-implicit Williamson algorithm, we may put these equations into the form in which only the adjustment terms appear on the left:

$$\psi^1_{\rm adj} = (1 - W^1_1 J^*)^{-1} W^1_0 F^0 - E^0, \tag{4.32}$$

$$\psi^2_{\rm adj} = -E^1, \tag{4.33}$$

$$\psi^3_{\rm adj} = (1 - W^3_3 J^*)^{-1} (W_E E^2 + W^3_2 F^2) - E^2, \tag{4.34}$$

$$\psi^4_{\rm adj} = -E^3, \tag{4.35}$$

where,

$$W^1_0 = \frac{1}{2}, \tag{4.36}$$

$$W^1_1 = \frac{1 + a_1}{4}, \tag{4.37}$$

$$W_E = -\frac{(1+\sqrt{2})b}{4}, \tag{4.38}$$

$$W^3_2 = \frac{1}{2} + \frac{(1+\sqrt{2})b}{8}, \tag{4.39}$$

$$W^3_3 = \frac{(1 + a_3 + \frac{1}{2}b)}{4}. \tag{4.40}$$

As we did for the third-order scheme, we can test the implicit modifications of the fourth-order Runge-Kutta scheme on the complex oscillator equation (2.1). Fig. 9 shows the stability curve (dotted, and symmetric about the real axis) for the unaltered explicit fourth-order scheme. The asymmetric curves roughly correspond to those shown in Fig. 4, as the implicit steps are pure trapezoidal without de-centering. As before, the implicit modes lack the property of robustness, which is only gained significantly for the faster implicit modes when $b > 0$, as shown in Fig. 10 (where $b = .5$) and Fig. 11 (where $b = 1$). All modes are rendered robust by choosing positive parameters a, as shown by Fig. 12 and, as with the Williamson modified scheme, a compromise of robustness and formal accuracy is obtained with a smaller value of 0.2 for a_1 and a_3 but a positive and b to guarantee the substantial robustness (or stronger damping) of the faster modes, as we see from Fig. 13. However, as before, the value of having a non-vanishing parameter b here is highly doubtful, especially in view of the significant simplifications to Eqs. (4.38)–(4.40) we gain and the resulting greater uniformity of treatment of implicit terms during both halves of the overall time stepping cycle when we choose $b = 0$.

The semi-implicit stages 2 and 4 for the Gill scheme look rather trivial but, in practice, unless a spectral projection method is used, these stages will still necessitate solution of elliptic equations. These now become equations of Poisson-type for the gravity or acoustic waves that we would typically wish to treat implicitly. For the case of gravity modes, the implied constraints are essentially equivalent to geostrophy in the corrections made at these even-indexed stages; for the case of acoustic modes, the implied constraints are essentially anelastic. In practice, the long range of influence that such constraints imply become problematic for numerical models spannning several processors of a massively-parallel computer. Because of such practical considerations, we would prefer that, in addition to the form of numerical robustness

we have described and discussed in the case of the semi-implicit Williamson modification, the semi-implicit Gill schemes be robust to a 'dilution' of the adjustments applied to the slower members of the designated 'fast' set of modes. That is, we desire that our schemes remain stable when the *actual* adjustments $\psi^*_{\rm adj}$ are derived from the *formal* adjustments $\psi_{\rm adj}$ by a dilution or under-relaxation:

$$\psi^*_{\rm adj} = q\psi_{\rm adj}, \tag{4.41}$$

for some mode-dependent $q \leq 1$. Even without recourse to the spectral projection of the modes concerned, it is possible to obtain such a q automatically by converting the original Poisson-type equation to one of Helmholtz-type by the appropriate choice of Helmholtz coefficient (which leaves the smallest scales essentially unaffected). In this way, the large-scale influence that is so troublesome on a parallel computer is effectively eliminated. Of course, if we *do* have convenient recourse to spectral projection, the dependence of dilution factor q on the modal frequency can be engineered with considerable freedom.

The dilution of adjustments should only take effect for the slowest of the implicit members of the family of modes (that is, large scale gravity modes in a hydrostatic model, large scale acoustic modes in a nonhydrostatic model). As we see from Fig 9, the stable region of the imaginary J axis is actually more inclusive for the the unmodified (explicit) Runge-Kutta scheme than it is for the modified (implicit) adaptation in the case of the slowest modes, so we might expect a dilution of the adjustment at all stages to be beneficial. Fig. 14 shows a succession of neutral stability curves for the single mode, $J^* = i$, for the modified Gill scheme with parameters $a_1 = a_3 = 0$, $b = 1$, (the scheme depicted in Fig. 11) for dilution factors, $q \in \{1.0, 0.75, 0.5, 0.25, 0.\}$. The scale of $\mathrm{Re}(J)$ is greatly expanded to facilitate visual discrimination of these curves relative to the imaginary axis (the curve for $q = 0$, which implies the unmodified explicit scheme, is the curve blending smoothly with the imaginary axis near the origin). We see from this sequence that, with progressive dilution, the stable segment of the imaginary J axis first actually shrinks slightly before it suddenly expands to the full span of the stable segment of the explicit scheme. This strategy therefore does not have the intended consequence of recovering robustness for the slower implicit modes for this particular modified scheme. The robustness for all modes that is achieved with the choice of $a_1 = a_3 = .5$ but $b = 0$, illustrated in Fig. 12, is preserved undiminished for the mode $J^* = i$ when progressively diluted, as we see in Fig. 15 (which adopts the same scale and format as Fig. 14). The dilution of the scheme, $a_1 = a_3 = .2$, $b = .5$, which attempted to compromise between accuracy and robustness (see Fig. 13) slightly degrades the robustness for all but the smallest dilution factors, q, as we see in Fig. 16, but probably not to the extent of jeopardizing the scheme.

We have shown what the effects are of diluting the adjustments by a uniform factor at all four stages but one of the reasons for considering this is to simulate the effect of converting the Poisson-type equations (implied by the undiluted adjustments of stages 2 and 4) to Helmholtz-type equations, so that the effective horizontal scale of influence of the terms that force these adjustments is reduced to an extent that is more manageable numerically in a parallel implementation. It is therefore instructive to see the effect that the dilution has when applied *only* to stages 2 and 4. The result, again for dilution factors progressing in intervals of $1/4$, is shown for the case $a_1 = a_3 = .5$ and $b = 0$ in Fig. 17. Here, it is the smaller values of q that begin to encroach on the region of oscillatory stability of the scheme, but not to the extent of seriously

degrading the overall robustness. However, we now lose the advantages of reverting to the classical Runge-Kutta solution in the limit $q \to 0$, especially the higher-order of accuracy, but also the substantial robustness of the Runge-Kutta scheme as applied to the slowest modes.

5. Discussion

The third-order and fourth-order Runge-Kutta schemes of Williamson (1980) and Gill (1951) have been shown to be amenable to semi-implicit modifications that extend their applicability to systems of equations that contain fast modes of no direct importance to the solution components of interest. Such systems include the hydrostatic and nonhydrostatic systems of equations employed by atmospheric forecasting and simulation models which have motivated this study, but other application might include ocean models, and models of the upper-terrestrial or solar atmospheres where electrical currents and their magnetic effects cannot be neglected.

Our study examined the robustness of several combinations of the first-order and second-order de-centering parameters. On the basis of the results we obtained for the linear damped oscillator equation, we cannot recommend attempting to achieve robustness using purely a second-order de-centering (via parameter b alone, in our notation), as the robustness attained falls far short of what is required for the slower members of the implicitly treated set of modes. One 'remedy', which we have avoided here, would be to artificially augment the frequencies of all the standard linearized implicit modes, that is, forcing our J^* to be always significantly larger than any actual effective J that could arise in the fully nonlinear system. For example, in the context of an atmospheric model, whether for acoustic or gravity modes, this effect could be accomplished by linearizing these modes about an artificial atmospheric basic state warmer and more statically stable than any realistic state likely to be encountered in a simulation. This is a device that is often used in semi-implicit atmospheric models, but we have avoided it because the error incurred seems less amenable to formal evaluation and control than is the error involved in our recommended remedy, which is to adopt a first-order decentering in time, using parameters $a > 0$ in the notation of Section 4. When we recall that the modes *requiring* implicit treatment are exclusively modes of characteristic time scales shorter than the time scales of interest for the simulation, we can justify the severe distortion of these modes that the implicit treatments imply. However, in recognizing that such modes occur in families, each spanning a spectrum of characteristic frequencies whose range may straddle the boundary between the 'important' (slow) and 'unimportant' (fast) time scales, we pay some attention to the obligation of avoiding the distortion of the evolution of the slower members of such a family more than can be helped. For the gravity modes, this means projecting only the deepest vertical modes for implicit treatment, as originally suggested by Burridge (1975). Among the deep subset of the gravity modes this might also include attempting to dilute the implicit adjustments at the largest horizontal scales (where frequencies become low enough to be stably resolved by the explicit time integration) by, for example, artifically increasing the Helmholtz coefficient of the Helmholtz elliptic problem beyond what is implied by a formal interpretation of the (undiluted) semi-implicit adjustment steps. This would have the effect of reducing the horizontal range of influence, especially in the case of the deepest family of modes (the Lamb modes). Likewise, for the acoustic modes of a nonhydrostatic model, it is *not* necessary to treat those modes of large

horizontal scales with a formally accurate representation of the horizontal part of the (three-dimensional) Helmholtz operator, as long as the usually dominant vertical part is properly dealt with; this allows the horizontal scale of influence in the treatment of these modes to be made much smaller, probably with little discernable detriment to the overall solution, by appropriately adjusting the Helmholtz coefficient in the elliptic problem or by directly diluting the adjustment in the spectral domain when this becomes feasible.

One aim of this study has been to identify methods that are robust to both errors in the modal frequencies (of the implicit modes) and to the effects of 'diluting' the implicit adjustment increments at the large scales (and hence, the slower modes) so that efficient, but accurate, implementations on massively-parallel platforms can be achieved for complete three-dimensional atmospheric models. In Part II of this study, we shall explore the practical aspects of implementation of semi-implicit modifications to the Runge-Kutta schemes in Eulerian and semi-Lagrangian nonlinear models. This will be done primarily in the context of the 'shallow water' system, which is sufficient to elucidate most of the potential hazards involved in applications of semi-implicit techniques to realistic simulation models. But, using a multi-level model, we shall also give consideration to some of the potential problems that can arise in applying semi-implicit techniques to the acoustic modes of a nonhydrostatic model, especially in the presence of non-trivial terrain. These more complete models will employ the first-order de-centering option that seems, from this Part I of the study, to offer the best combination of robustness and simplicity.

Acknowledgments

This paper benefitted from the helpful comments of Drs. S. G. Gopalakrishnan and J. Sela. This work was partially supported by the NSF/NOAA Joint Grants Program of the US Weather Research Program. This research is also in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

Appendix A

One-step and Runge-Kutta schemes

The explicit schemes in this category involve a cycle of n stages and may be defined, for a generic cycle of duration δt as follows:

$$\psi^0 = \psi(t), \tag{A.1}$$

$$F^j = \mathcal{F}(\psi^j)\delta t, \tag{A.2}$$

$$\psi^j = \psi^0 + \sum_{i=0}^{j-1} a_{j,i} F^i, \qquad j = 1, n, \tag{A.3}$$

$$\psi(t+\delta t) = \psi^n. \tag{A.4}$$

The subdivisions of δt at which the intermediate function evaluations occur are $c_j \delta t$, where,

$$c_j = \sum_{i=0}^{j-1} a_{j,i}, \tag{A.5}$$

where $c_n = 1$ always. A convenient tableau for defining the coefficients of, for example, a three-stage scheme of this form, is:

c_1	$a_{1,0}$		
c_2	$a_{2,0}$	$a_{2,1}$	
	$a_{3,0}$	$a_{3,1}$	$a_{3,2}$

(A.6)

For such a scheme to belong to the family of three-stage third-order Runge-Kutta methods, Butcher (1987) shows that:

$$a_{3,0} + a_{3,1} + a_{3,2} = 1, \tag{A.7}$$

$$a_{3,1}c_1 + a_{3,2}c_2 = \frac{1}{2}, \tag{A.8}$$

$$a_{3,1}c_1^2 + a_{3,2}c_2^2 = \frac{1}{3}, \tag{A.9}$$

$$a_{3,2}a_{2,1}c_1 = \frac{1}{6}. \tag{A.10}$$

The two families of the N-cycle methods of Lorenz (1971) are defined as the special cases of one-step methods in which the state at subsequent stages is constructed cumulatively by schemes algebraically equivalent to:

$$\psi^j = \psi^{j-1} + \frac{1}{n}E^{j-1}, \qquad j = 1, n, \tag{A.11}$$

$$E^0 = F^0, \tag{A.12}$$

where, for scheme 1:

$$E^j = \frac{n}{n-j}F^j - \frac{j}{n-j}E^{j-1}, \qquad j = 1, n-1, \tag{A.13}$$

and, for scheme 2:

$$E^j = \frac{n}{j}F^j - \frac{n-j}{j}E^{j-1}, \qquad j = 1, n-1. \tag{A.14}$$

The equivalent tableaux for the two 3-cycle schemes are:

$\frac{1}{3}$	$\frac{1}{3}$		
$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	
	$\frac{1}{2}$	$-\frac{1}{2}$	1

$\frac{1}{3}$	$\frac{1}{3}$		
$\frac{2}{3}$	$-\frac{2}{3}$	1	
	0	$\frac{1}{2}$	$\frac{1}{2}$

Both schemes fail to satisfy (A.9) (with equal and opposite errors) and are therefore not strictly fully third-order accurate except in the case of a linear relationship between the ψ and the F. The same limitations apply to all Lorenz N-cycle schemes for $N > 2$.

The schemes investigated by Williamson (1980) are the fully third-order schemes that are possible by generalizing Lorenz's 3-cycle method to intermediate time levels that do *not* equally subdivide the cycle, δt. Apart from a few isolated degenerate cases, it can be shown, by a minor

modification of the argument of Williamson, that these schemes all correspond to an algebraic curve,

$$Y^2(1-X+\frac{X^2}{3})+Y(-1+\frac{3}{2}X-X^2)+(-X+X^2)=0, \tag{A.15}$$

which is symmetric with respect to the interchange of X and Y, where,

$$X = 1/c_1, \tag{A.16}$$
$$Y = 1/(1-c_2), \tag{A.17}$$

are the reciprocals of the first and third time subdivisions. Fig. 18 illustrates this curve, together with a schematic depiction of a process for generating a two-way sequence of methods possessing rational coefficients. The idea is that, since (A.15) has rational coefficients and is separately quadratic in X and in Y, then, given any rational solution, $S_{2p}=(X_{2p},Y_{2p})$, there is a related rational solution, S_{2p+1}, possessing the same Y, and with X given by

$$X_{2p+1}=\frac{3(2-3Y_{2p}+2Y_{2p}^2)}{2(3-3Y_{2p}+Y_{2p}^2)}-X_{2p} \tag{A.18}$$

Interchanging X and Y, the corresponding formula allows us to derive the rational scheme S_{2p+2} from S_{2p+1}, and so on. The natural starting point for the inductive construction is $S_0\equiv(0,0)$. However, this point, together with $S_{-1}\equiv(0,1)$ and $S_1\equiv(1,0)$, do not translate into meaningful integration schemes owing to the implied indeterminacies in the interval subdivisions. Less obviously, $S_3\equiv(1/3,1/3)$ cannot be associated with any Runge-Kutta scheme with finite coefficients, but the remaining members of the double-sided sequence correspond to rational-coefficient low-storage Runge-Kutta schemes. This includes $S_4\equiv(3,4)$, which is *the* recommended Williamson scheme. Its tableau is,

$$\begin{array}{c|ccc} \frac{1}{3} & \frac{1}{3} & & \\ \frac{3}{4} & -\frac{3}{16} & \frac{15}{16} & \\ \hline & \frac{1}{6} & \frac{3}{10} & \frac{8}{15} \end{array} \tag{A.19}$$

Table 3 lists a few of the schemes in this sequence; while the sequence can be shown to sample densely the locus of possible low-storage schemes shown in Fig. 18, the convenience of having rational coefficients is offset by the fact that the rationals involved only possess small integer numerators and denominators for the few members S_p with small $|p|$. If there is no compelling reason to restrict oneself to such rational schemes, the unique scheme $\bar{S}$ for which $X=Y\neq 0$ becomes an attractive choice.

Expanding the terms in the algorithm of Table 1 and using (A.7)–(A.10) we find the coefficients of the generic Williamson scheme, given c_1 and c_2, from:

$$a_{3,1} = \frac{3c_2-2}{6c_1(c_2-c_1)}, \tag{A.20}$$
$$a_{3,2} = \frac{2-3c_1}{6c_2(c_2-c_1)}, \tag{A.21}$$

TABLE 3. SOME LOW-STORAGE RK3 SCHEMES WITH RATIONAL COEFFICIENTS

Scheme	Index used by Williamson	X	Y	c_1	c_2	R_0	R_1	R_2	Q_1	Q_2
S_{-5}	5	4	$\frac{12}{7}$	$\frac{1}{4}$	$\frac{5}{12}$	$\frac{1}{4}$	$\frac{2}{9}$	3	$-\frac{2}{9}$	$-\frac{29}{2}$
S_{-4}	6	4	3	$\frac{1}{4}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{8}{9}$	$\frac{3}{4}$	$-\frac{17}{9}$	-1
S_{-3}	12	$\frac{3}{2}$	3	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{1}{3}$	$-\frac{9}{8}$	$-\frac{4}{9}$
S_{-2}	4	$\frac{3}{2}$	1	$\frac{2}{3}$	0	$\frac{2}{3}$	$-\frac{3}{4}$	$-\frac{1}{3}$	$\frac{1}{8}$	-2
S_2	14	1	$\frac{3}{2}$	1	$\frac{1}{3}$	1	$\frac{2}{9}$	$\frac{3}{4}$	$-\frac{8}{9}$	$\frac{1}{8}$
S_4	7	3	4	$\frac{1}{3}$	$\frac{3}{4}$	$\frac{1}{3}$	$\frac{15}{16}$	$\frac{8}{15}$	$-\frac{25}{16}$	$-\frac{17}{25}$
S_5	...	$\frac{12}{7}$	4	$\frac{7}{12}$	$\frac{3}{4}$	$\frac{7}{12}$	$\frac{6}{7}$	$\frac{1}{3}$	$-\frac{58}{49}$	$-\frac{1}{3}$

$$R_0 = c_1, \tag{A.22}$$
$$R_2 = a_{3,2}, \tag{A.23}$$
$$R_1 = \frac{1}{6R_0R_2}, \tag{A.24}$$
$$Q_1 = \frac{c_2 - c_1 - R_1}{R_0}, \tag{A.25}$$
$$Q_2 = \frac{a_{3,1}}{R_1} - 1. \tag{A.26}$$

For the symmetric scheme, $\bar{S}$, we find that

$$\begin{aligned} c_1 &\approx 0.28771294386878, \\ c_2 &\approx 0.71228705613122, \\ R_0 &\approx 0.28771294386878, \\ R_1 &\approx 0.92457411226239, \\ R_2 &\approx 0.62653829327082, \\ Q_1 &\approx -0.5407895304104, \\ Q_2 &\approx -1.1776506988040. \end{aligned}$$

The tableau of coefficients generalizes in the obvious way for fourth-order schemes. The classical fourth-order Runge-Kutta scheme is defined by its tableau,

$$\begin{array}{c|cccc} \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \tag{A.27}$$

For the Gill scheme, the corresponding coefficients are:

$$\begin{array}{c|ccc} \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & -\frac{1}{2}+\sqrt{\frac{1}{2}} & 1-\sqrt{\frac{1}{2}} & \\ 1 & 0 & -\sqrt{\frac{1}{2}} & 1+\sqrt{\frac{1}{2}} \\ \hline & \frac{1}{6} & \frac{1}{3}\left(1-\sqrt{\frac{1}{2}}\right) & \frac{1}{3}\left(1+\sqrt{\frac{1}{2}}\right) \quad \frac{1}{6} \end{array} \tag{A.28}$$

Appendix B

Delineation of stability boundaries for idealized complex oscillator equations

First, consider the case of the third-order Adams-Bashforth (AB3) scheme and suppose the forcing F to be a linear function of the state ψ, both being represented by complex magnitudes. The complex frequency J is simply a constant of proportionality in the linear relationship,

$$F = J\psi,$$

while, for any single characteristic mode (physical or computational) of the AB3 scheme with time step δt, the state evolution obeys

$$\psi(t + m\delta t) = \nu^m \psi(t), \tag{B.1}$$

for the *numerical* complex frequency characterizing this mode. Clearly, the region of the plane of parameter J corresponding to numerical stability also corresponds to where the values ν do not exceed unity in absolute magnitude. The stable region is therefore bounded by the curve, or curves, where

$$\nu(J) = \exp(i\theta), \tag{B.2}$$

for real angles θ. Recalling the form of the AB3 scheme:

$$\psi(t+\delta t) = \psi(t) + \frac{1}{12}\left(23F(t) - 16F(t-\delta t) + 5F(t-2\delta t)\right), \tag{B.3}$$

the substitutions for a single mode lead immediately to

$$\nu\psi(t) = \left(1 + \frac{J}{12}(23 - 16/\nu + 5/\nu^2)\right)\psi(t). \tag{B.4}$$

Therefore,

$$J(\theta) = \frac{12\nu^2(\nu-1)}{23\nu^2 - 16\nu + 5}, \tag{B.5}$$

with (B.2) makes it possible to plot this bounding curve of $J(\theta)$ as a function of all values of θ from 0 to 2π, as is done in Fig. 1. In a similar manner, the stability boundary for the time-filtered leapfrog method of Fig. 1, which is also a multistep method, is easily obtained.

For the classical fourth-order Runge-Kutta method defined by the tableau of (A.27) the sequence of stages $\psi^1, \ldots \psi^4$, expanded in terms of $\psi^0 \equiv \psi(t)$ is:

$$\psi^0 \equiv \psi(t), \tag{B.6}$$
$$\psi^1 = (1 + J/2)\psi^0, \tag{B.7}$$
$$\psi^2 = (1 + J/2 + J^2/4)\psi^0, \tag{B.8}$$
$$\psi^3 = (1 + J + J^2/2 + J^3/4)\psi^0, \tag{B.9}$$
$$\psi(t + \delta t) \equiv \psi^4 = (1 + J + J^2/2 + J^3/6 + J^4/24)\psi^0. \tag{B.10}$$

In this case, solving for the critically stable $J(\theta)$ when $\psi^4 = \exp(i\theta)\psi^0$ entails finding the solution of a quadratic for each value of θ. Standard polynomial root-finding software, such as is found in Press et al. (1989), is used for this task. Achieving continuity of the plots of the four roots $J(\theta)$ as angle θ makes a complete cycle requires careful selection of the appropriate permutation of these roots at each consecutive evaluation. The Gill (1951) scheme, with the standard tableau representation given by (A.28), leads to precisely the same final polynomial (B.10) as the classical scheme (A.27). The corresponding polynomial for the Nth-order Runge-Kutta scheme for $N \leq 4$ or for the general N-cycle schemes featured in Fig. 2 is the Nth-degree Taylor series approximation to the function $\exp(J)$, giving N independent complex roots. The semi-implicit adaptations we have described alter the polynomial coefficients but, except in degenerate limiting cases, not the degree. The same root-finding and permutation-matching procedure therefore allows us to construct the stability boundaries we show in the other figures of this paper.

References

Asselin, R. A. 1972 Frequency filter for time integrations. *Mon. Wea. Rev.*, **100,** 487–490.

Bates, J. R., and A. McDonald 1982 Multiply-upstream, semi-Lagrangian advective schemes: analysis and application to a multi-layer primitive equation model. *Mon. Wea. Rev.*, **110,** 1831–1842.

Burridge, D. M. 1975 A split semi-implicit reformulation of the Bushby-Timpson 10 level model. *Quart. J. Roy. Meteor. Soc.*, **101,** 777–792.

Butcher, J. C. 1965 On the attainable order of Runge-Kutta methods. *Math. Comp.*, **19,** 408–417.

Butcher, J. C. 1987 *The Numerical Analysis of Ordinary Differential Equations.* John Wiley, New York. 512pp.

Durran, D. R. 1991 The third-order Adams-Bashforth method: An attractive alternative to the leapfrog time-differencing scheme. *Mon. Wea. Rev.*, **119,** 702–720.

Durran, D. R. 1999 *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics.* Springer-Verlag, New York. 465pp.

Gear, C. W. 1971 *Numerical Initial Value Problems in Ordinary Differential Equations* Prentice-Hall, Englewood Cliffs, New Jersey, 253 pp.

Gill, S. 1951 A process for the step-by-step integration of differential equations in an automatic digital computing machine. *Proc. Camb. Philos. Soc.*, **47,** 96–108.

Haltiner, G. J., and R. T. Williams 1980 *Numerical Prediction and Dynamic Meteorology, 2nd Ed.* John Wiley, New York. 477pp.

Kopal, Z. 1961 *Numerical Analysis, 2nd ed.*, Chapman and Hall, London.

Kurihara, Y.	1965	On the use of implicit and iterative methods of time integration of the wave equation. *Mon. Wea. Rev.*, **93,** 33–46.
Lorenz, E. N.	1971	An N-cycle time-differencing scheme for stepwise numerical integration. *Mon. Wea. Rev.*, **99,** 644–648.
Magazenkov, L. N.	1980	Time integration schemes for fluid dynamics equations, effectively damping the high frequency components. *Tr. Gl. Geofiz. Observ.*, **410,** 120-129.
Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling	1989	*Numerical Recipes.* Cambridge University Press, Cambridge, U. K. 702 pp.
Purser, R. J., and L. M. Leslie	1988	A semi-implicit semi-Lagrangian finite-difference scheme using high-order spatial differencing on a nonstaggered grid. *Mon. Wea. Rev.*, **116,** 2067–2080.
Purser, R. J., and L. M. Leslie	1991	An efficient interpolation procedure for high-order three-dimensional semi-Lagrangian models. *Mon. Wea. Rev.*, **119,** 2492–2498.
Purser, R. J., and L. M. Leslie	1994	An efficient semi-Lagrangian scheme using third-order semi-implicit time integration and forward trajectories. *Mon. Wea. Rev.*, **122,** 745–756.
Purser, R. J., and L. M. Leslie	1996	Generalized Adams-Bashforth time integration schemes for a semi-Lagrangian model employing the second-derivative form of the horizontal momentum equations. *Quart. J. Roy. Meteor. Soc.*, **122,** 737–763.
Purser, R. J., and L. M. Leslie	1997	High-order generalized Lorenz N-cycle schemes for semi-Lagrangian methods employing second derivatives in time. *Mon. Wea. Rev.*, **125,** 1261–1276.
Robert, A. J.	1966	The integration of a low order spectral form of the primitive meteorological equations. *J. Meteor. Soc. Japan, Ser. 2,* **44,** 237–245.
Robert, A.	1981	A stable numerical integration scheme for the primitive meteorological equations. *Atmos. Ocean,* **19,** 35–46.
Robert, A.	1982	A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations. *J. Meteor. Soc. Japan,* **60,** 319–325.
Robert, A., J. Henderson and C. Turnbull	1972	An implicit time integration scheme for baroclinic models of the atmosphere. *Mon. Wea. Rev.*, **100,** 329–335.
Simmons, A. J., B. Hoskins, and D. Burridge	1978	Stability of the semi-implicit method of time integration. *Mon. Wea. Rev.*, **106,** 405–412.
Skamarock, W. C., J. B. Klemp, and J. Dudhia	2001	Prototypes for the WRF (Weather Research and Forecasting) model. *(Preprint) AMS 18th Conference on Weather Analysis and Forecasting and 14th Conference on Numerical Weather Prediction,* 30 July – 2 August 2001, Ft. Lauderdale, FL. pp. J11–J15.
Störmer, C.	1907	Sur les trajectoires des corpuscules électrisés. *Arch. Sci. Phys. Nat. Genève,* **24,** 5–18, 113–158, 221–247.
Wicker, L. J., and W. C. Skamarock	2001	Time splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev.*, (Submitted).
Williamson, J. H.	1980	Low storage Runge-Kutta schemes. *J. Comput. Phys.*, **35,** 48–56.

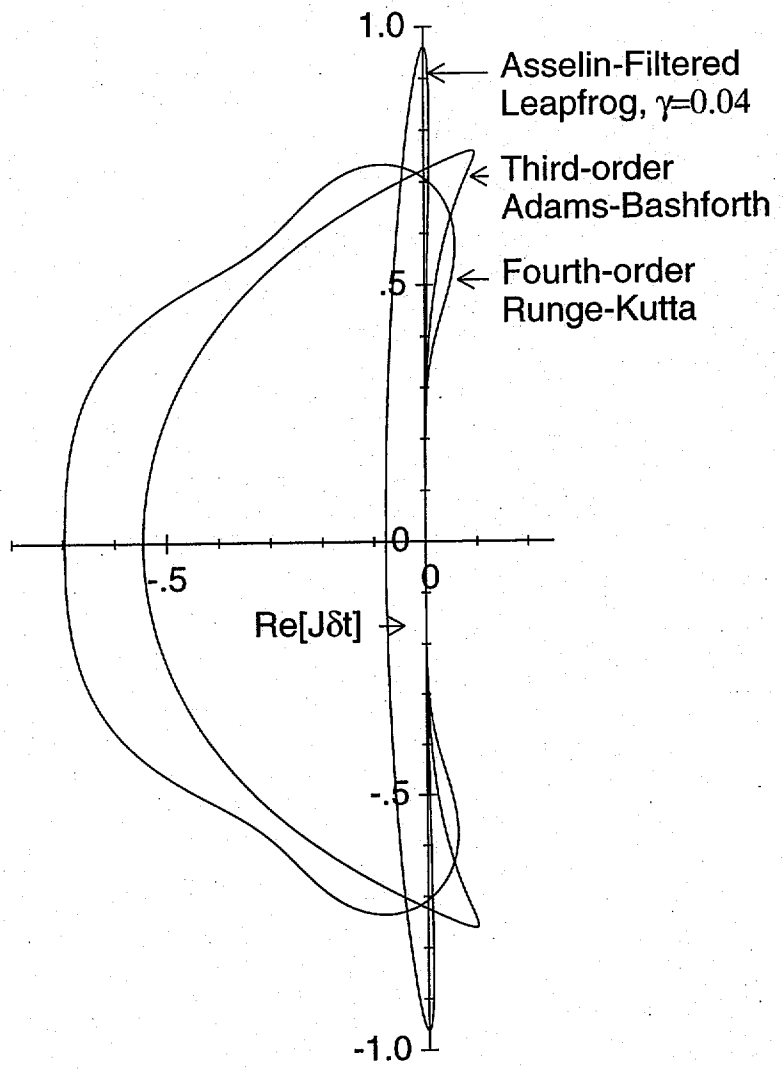

Figure 1. Stability regions in the complex $J\delta t$ plane of the filtered leapfrog, the third-order Adams-Bashforth, and the fourth-order Runge-Kutta schemes, all adjusted to possess the same average frequency of tendency function evaluations.

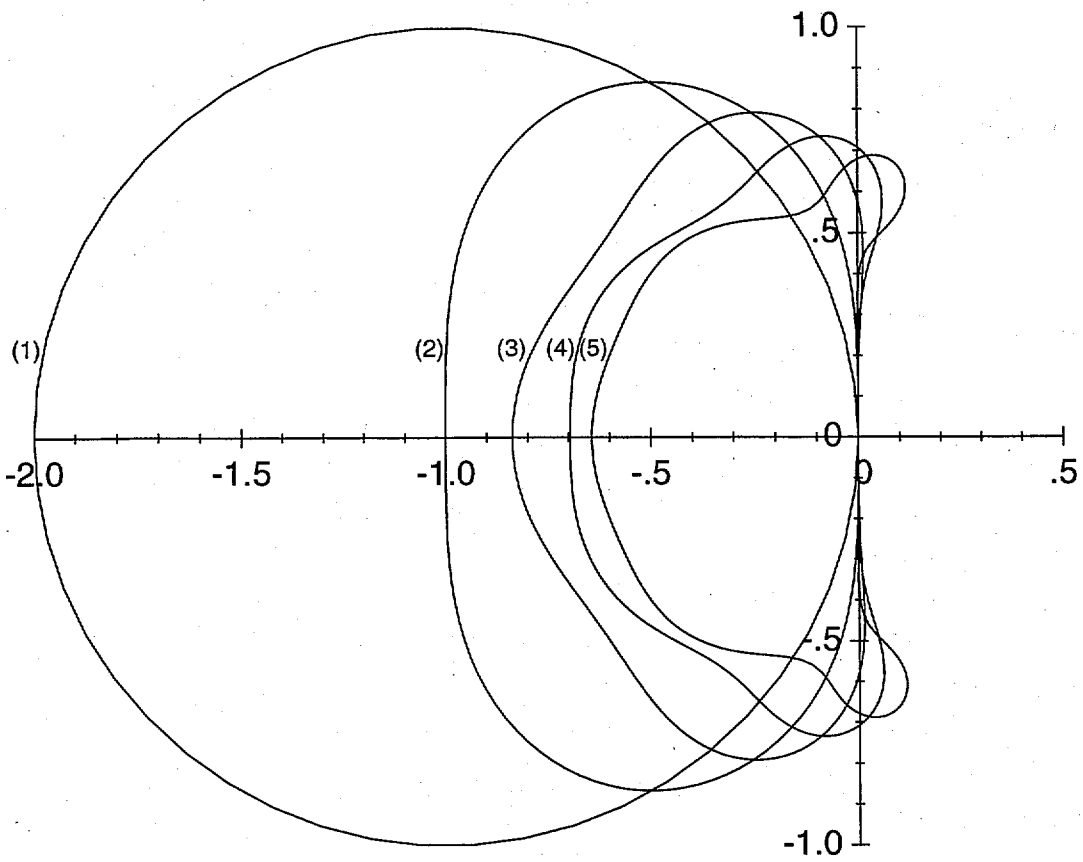

Figure 2. Stability regions in the complex $J\delta t$ plane of the N-cycle Runge-Kutta schemes up to $N = 5$, all adjusted to possess the same average frequency of tendency function evaluations.

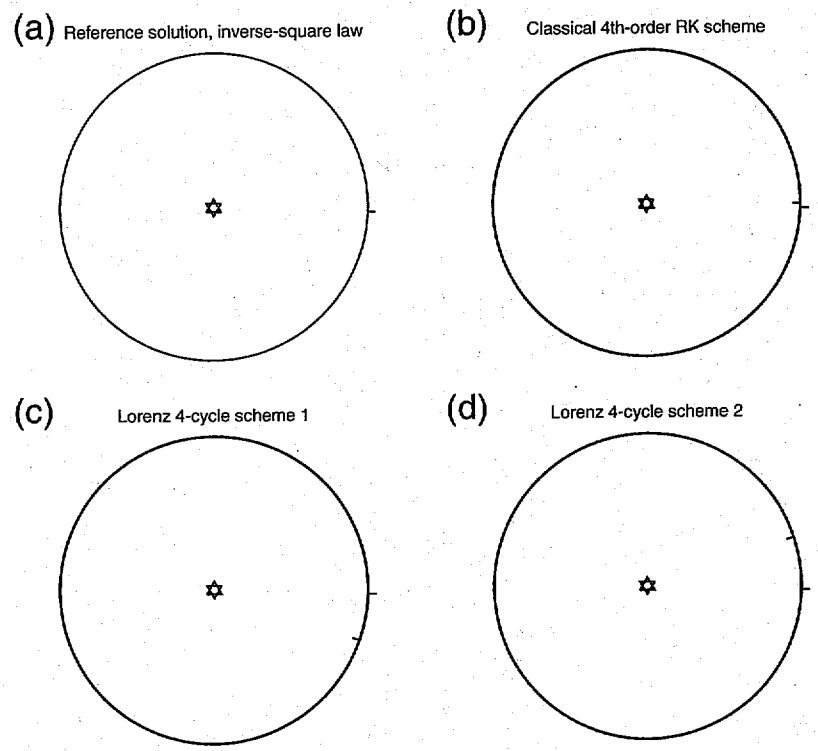

Figure 3. Numerical solution of the inverse-square force law problem, for approximately 10 circular orbits and 314 time-stepping cycles, comparing the classical 4th-order Runge-Kutta scheme with the two Lorenz 4-cycle methods. The true phase at the end of the period of integration is shown by the outer tick, the numerical phase for each scheme is indicated by the inner tick.

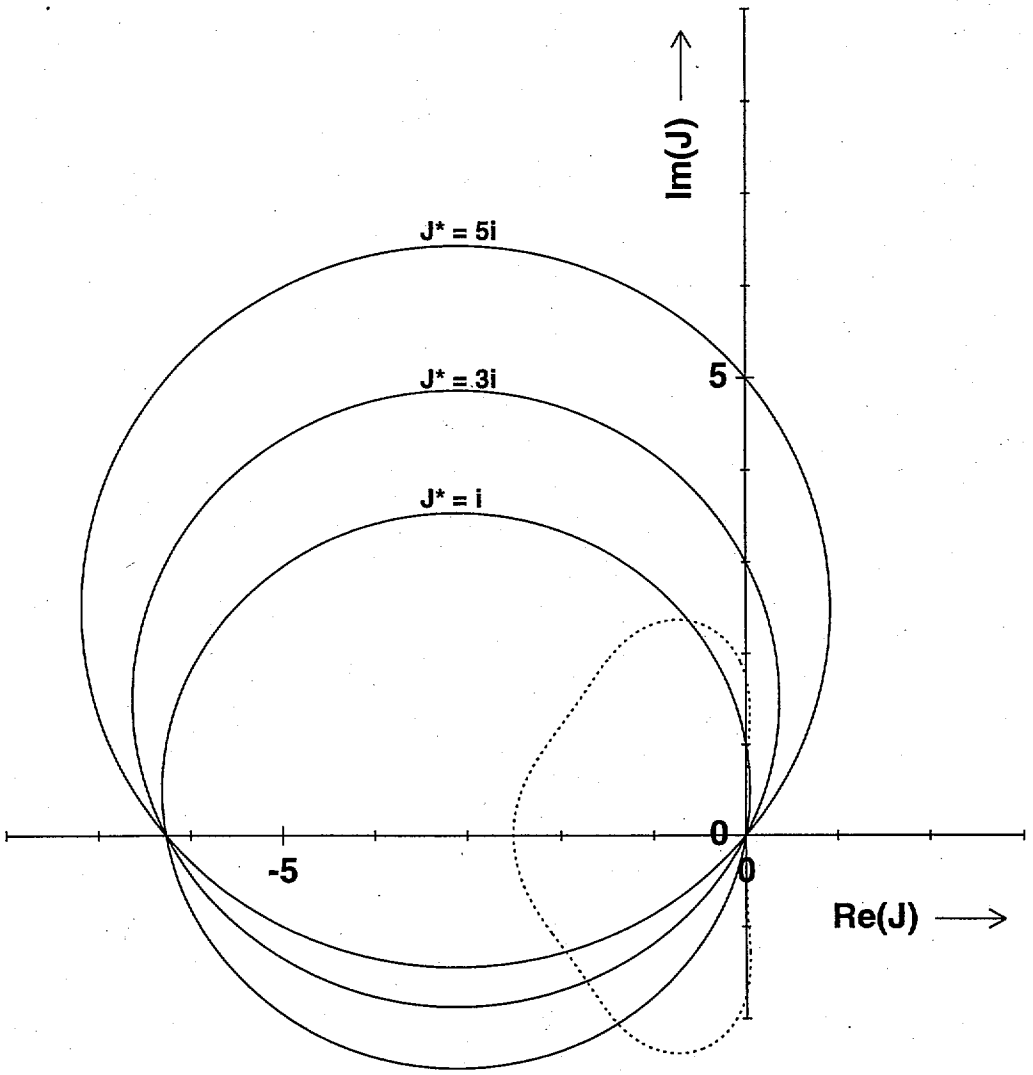

Figure 4. Stability regions (interiors of closed curves) in the complex plane of $J\delta t$ for the explicit third-order Runge-Kutta scheme (the dotted curve that is symmetric about the real-axis) together with the stability curves for the semi-implicit modification of Williamson's scheme at assumed frequencies, $J^* = i, 3i, 5i$ for the parameter choice, $a_1 = a_2 = a_3 = b = 0$. J is the true complex frequency of the mode. The curves for the implicit modified scheme all pass precisely through the respective complex values of J^*, implying that the fast oscillatory modes are exactly neutrally stable when $J^* = J$, but are unstable for the smallest underestimation, $J^* < J$, which implies a lack of numerical robustness.

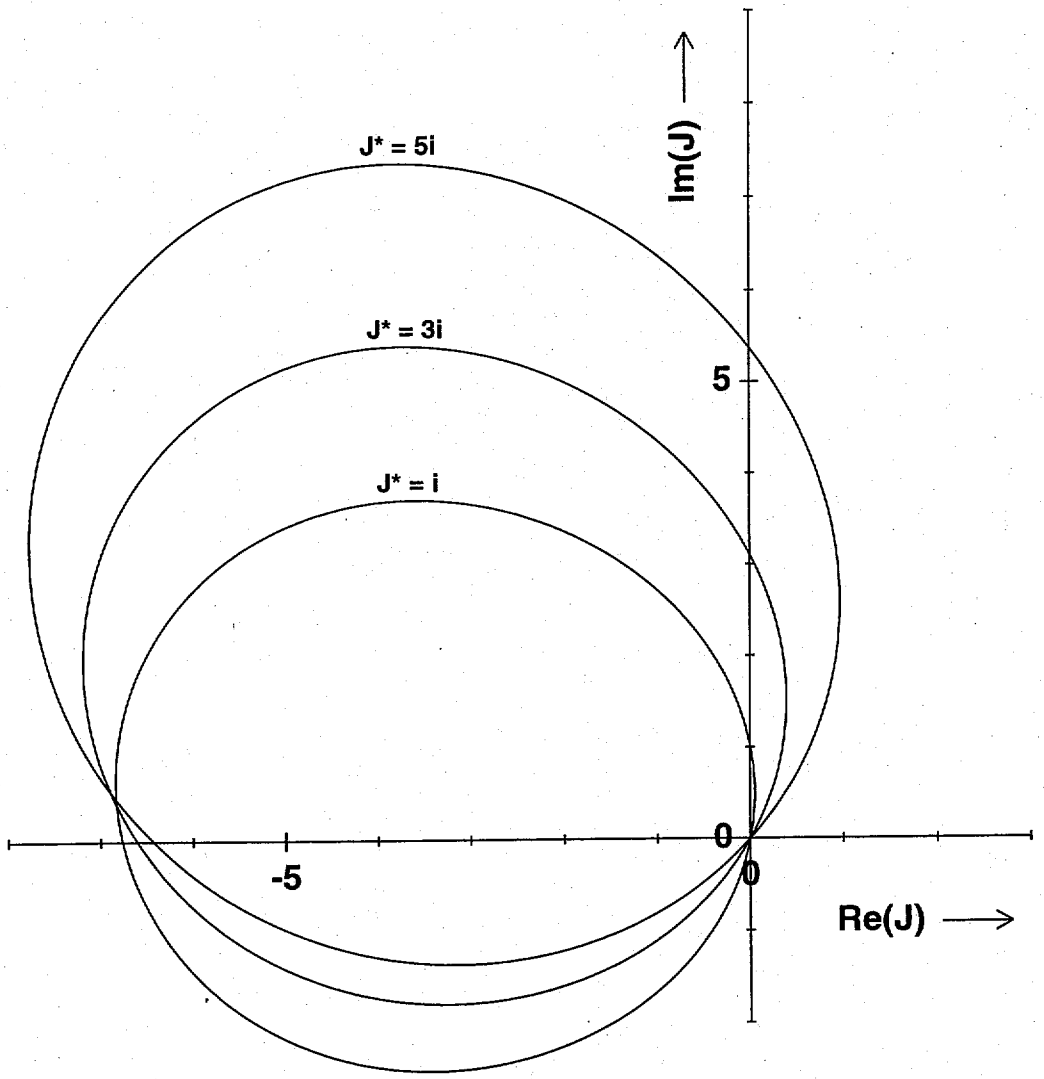

Figure 5. Stability regions in the J plane for the semi-implicit modified Williamson scheme for $J^* = i, 3i, 5i$ when $a_1 = a_2 = a_3 = 0$. but with the second-order de-centering parameter, $b = .5$. For the higher frequencies (larger loops), a degree of robustness is obtained, since the curve of neutral stability misses the point $J = J^*$ by a significant marging, but this margin is not significant at smaller frequencies (e.g., for the curve for $J^* = i$).

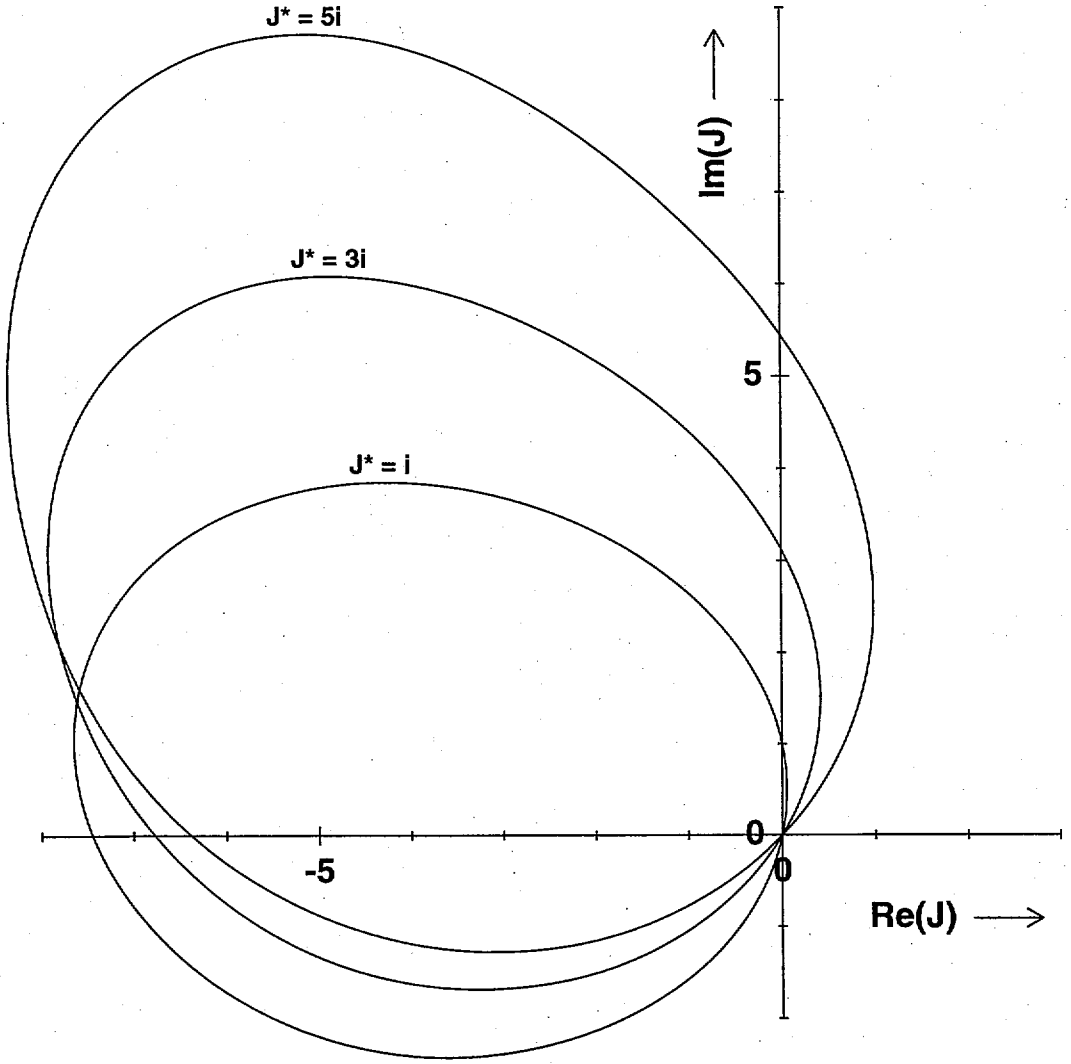

Figure 6. Modified Williamson scheme, as in Fig. 5, but for a larger de-centering parameter, $b = 1$. The lower frequency implicit modes still lack robustness.

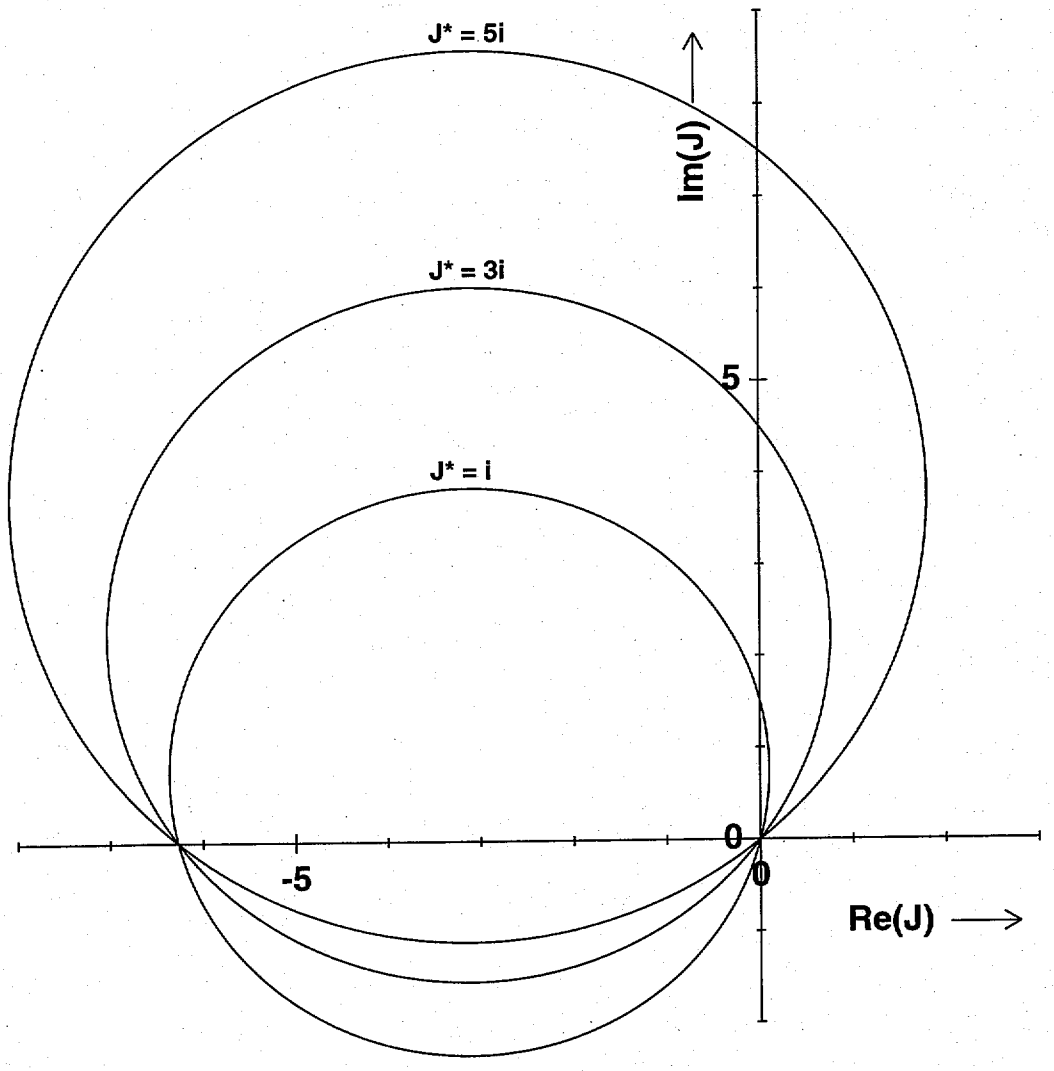

Figure 7. Modified Williamson scheme, as in Fig. 5, but with $a_1 = a_2 = a_3 = .5$ and $b = 0$.. All implicit modes are now robustly stable.

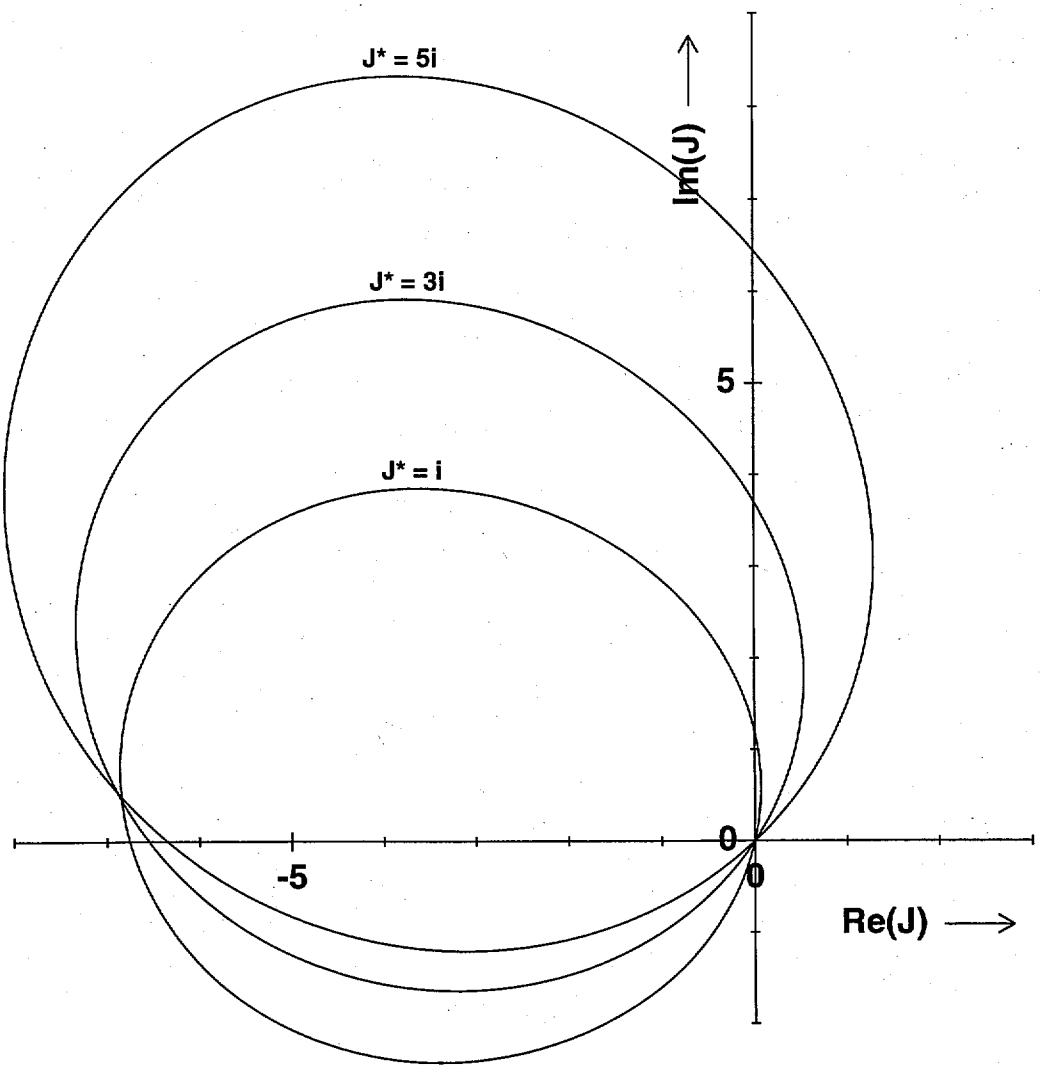

Figure 8. Modified Williamson scheme, as in Fig. 5, but with $a_1 = a_2 = a_3 = .2$ and $b = .5$. The small first-order de-centering parameters a provide some robustness for the lower frequencies while the b parameter provides a more generous safety margin for the higher frequencies.

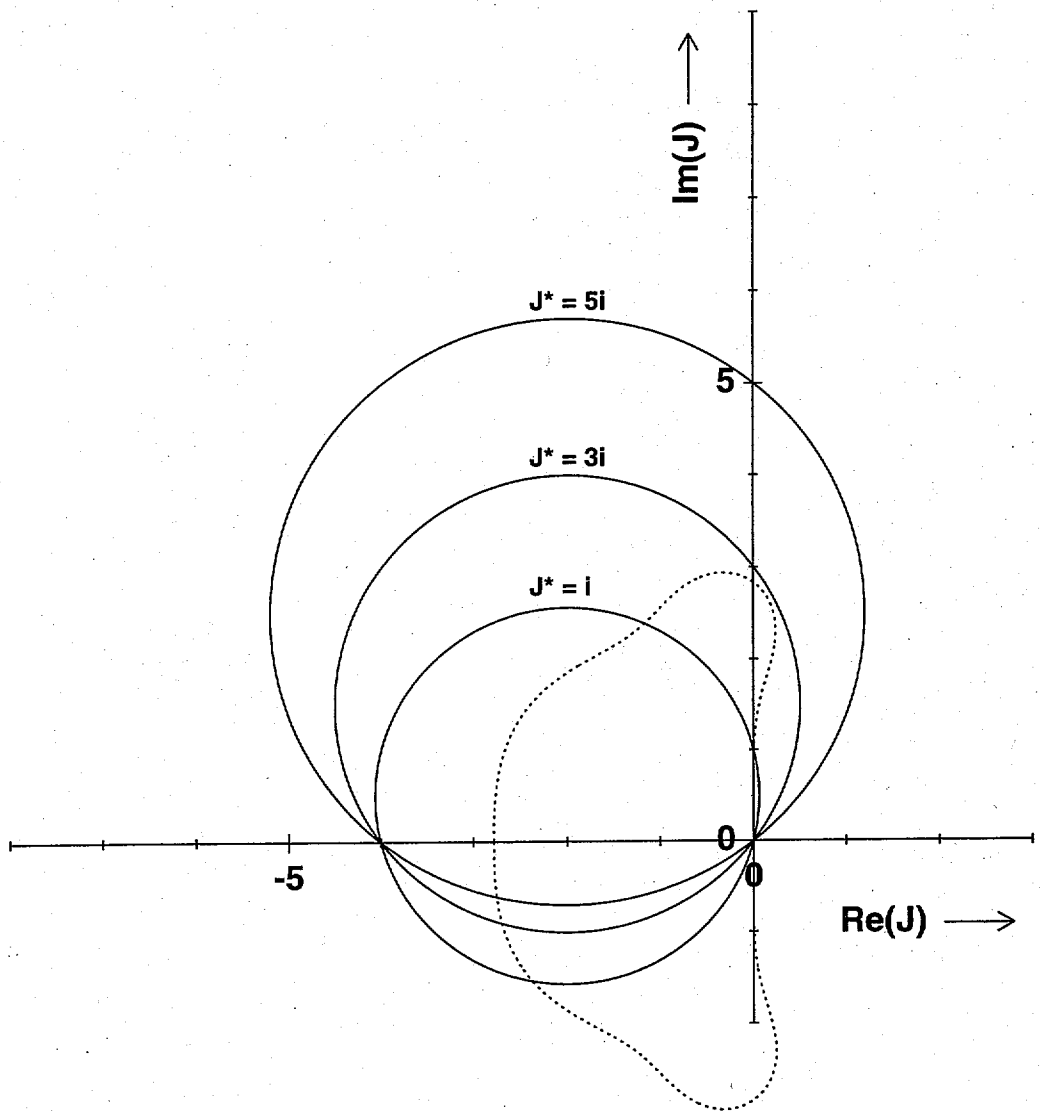

Figure 9. Stability regions as in Fig. 4, but for the explicit fourth-order Runge-Kutta scheme (the dotted curve that is symmetric about the real-axis) together with the stability curves for the semi-implicit modification of Gill's scheme at assumed frequencies, $J^* = i, 3i, 5i$ for the parameter choice, $a_1 = a_3 = b = 0$. As in the case of the modified Williamson scheme, whose stability curves are shown in Fig. 4, the implicit scheme's stability bounds all pass precisely through the respective complex values of J^*, implying an inherent lack of robustness at all frequencies.

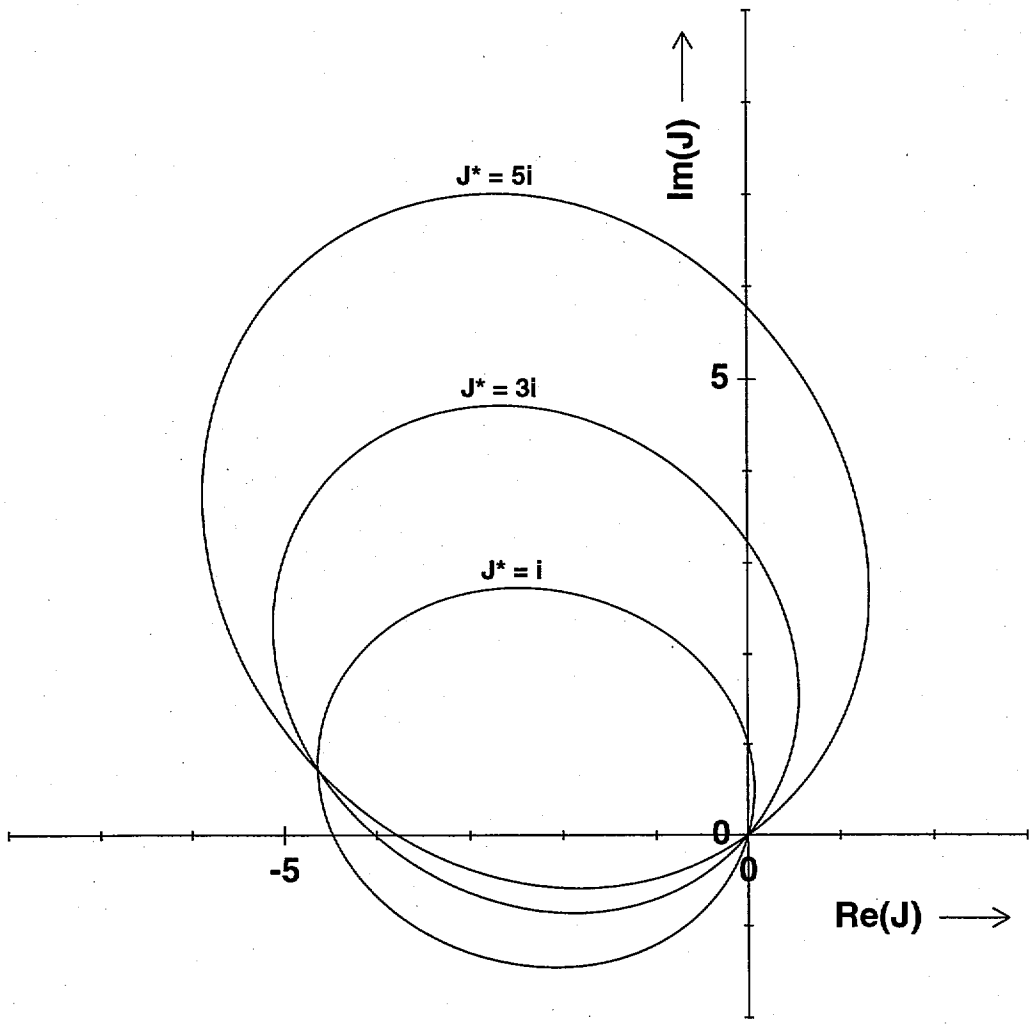

Figure 10. Stability regions in the J plane for the semi-implicit modified Gill scheme for $J^* = i, 3i, 5i$ when $a_1 = a_3 = 0$. but with the second-order de-centering parameter, $b = .5$. As with the corresponding modified Williamson scheme of Fig. 5, the higher frequencies (larger loops) are robust but the lower frequencies much less so.

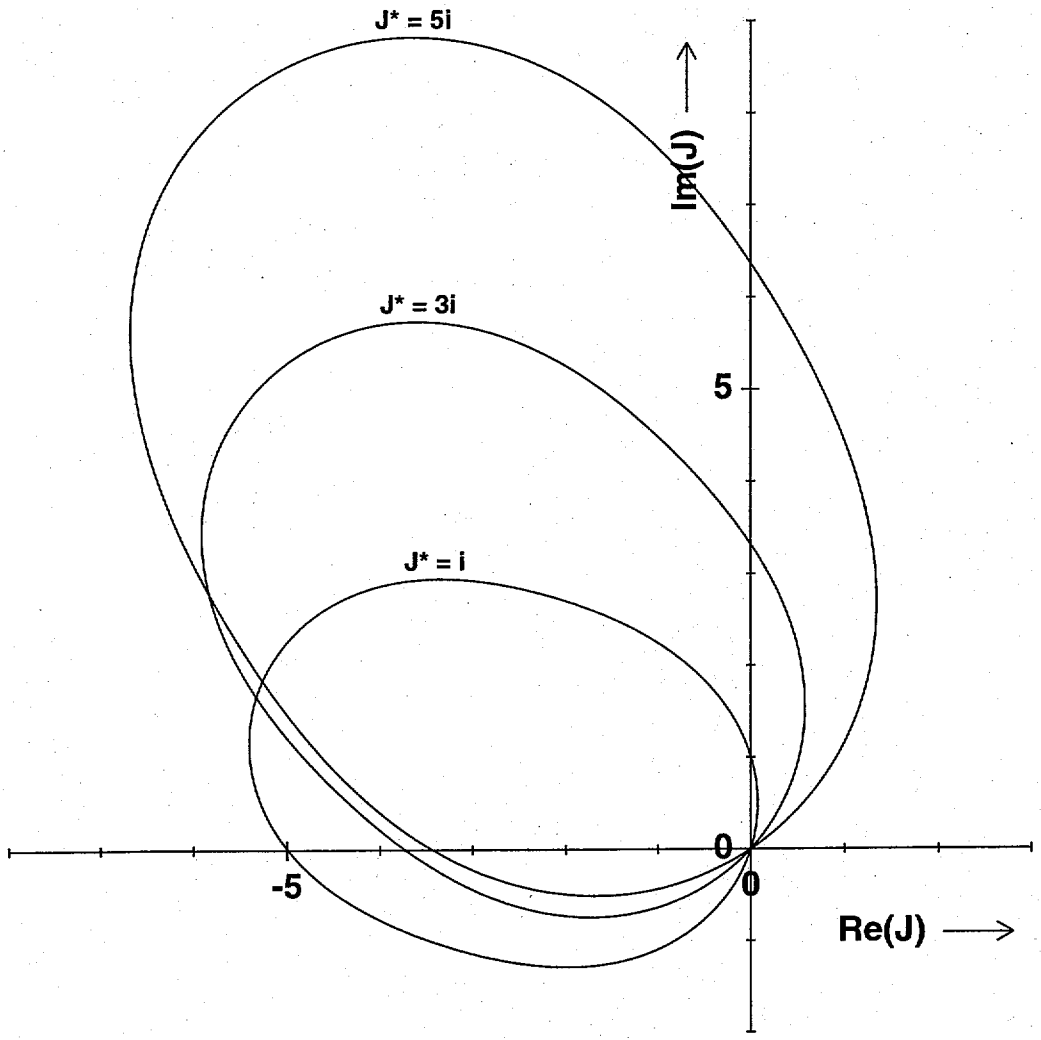

Figure 11. As in Fig. 10, but with $b = 1$.

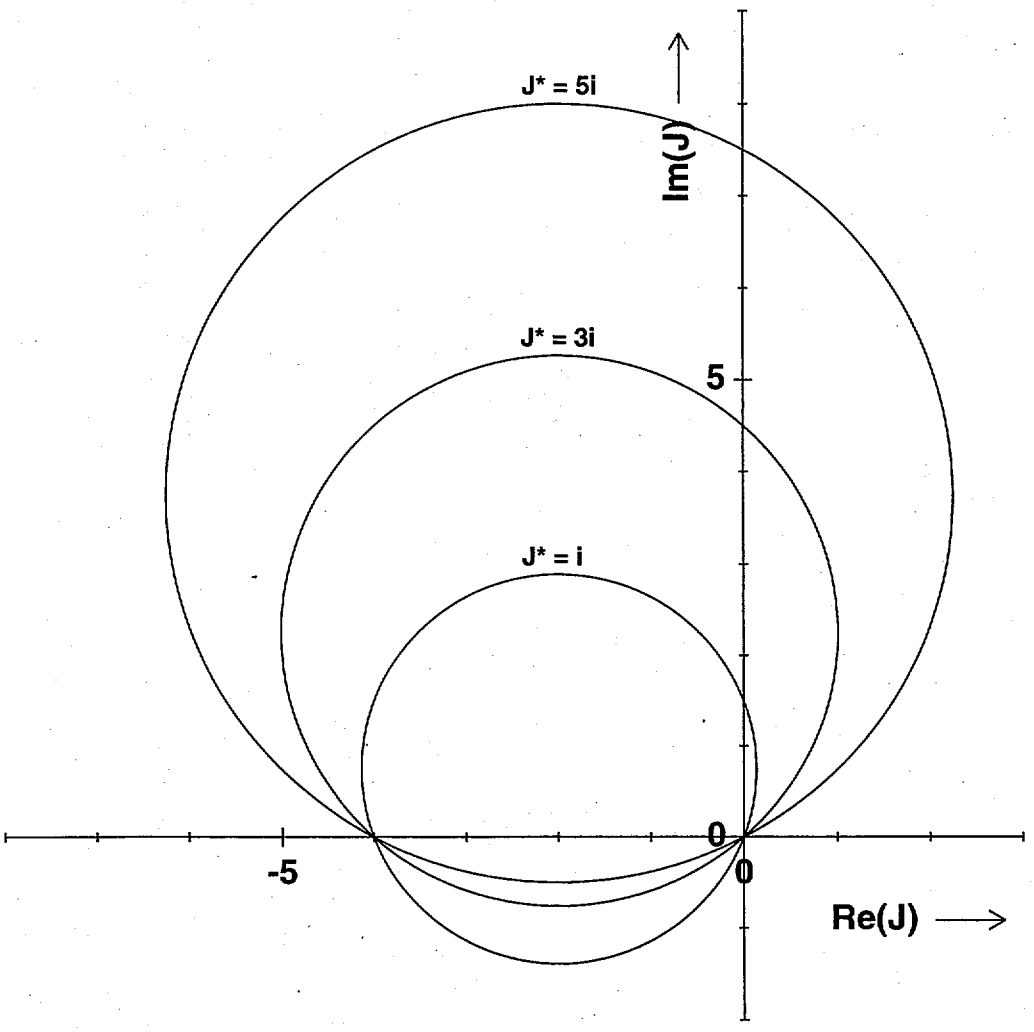

Figure 12. As in Fig. 10, but with $a_1 = a_3 = .5$ and $b = 0$.

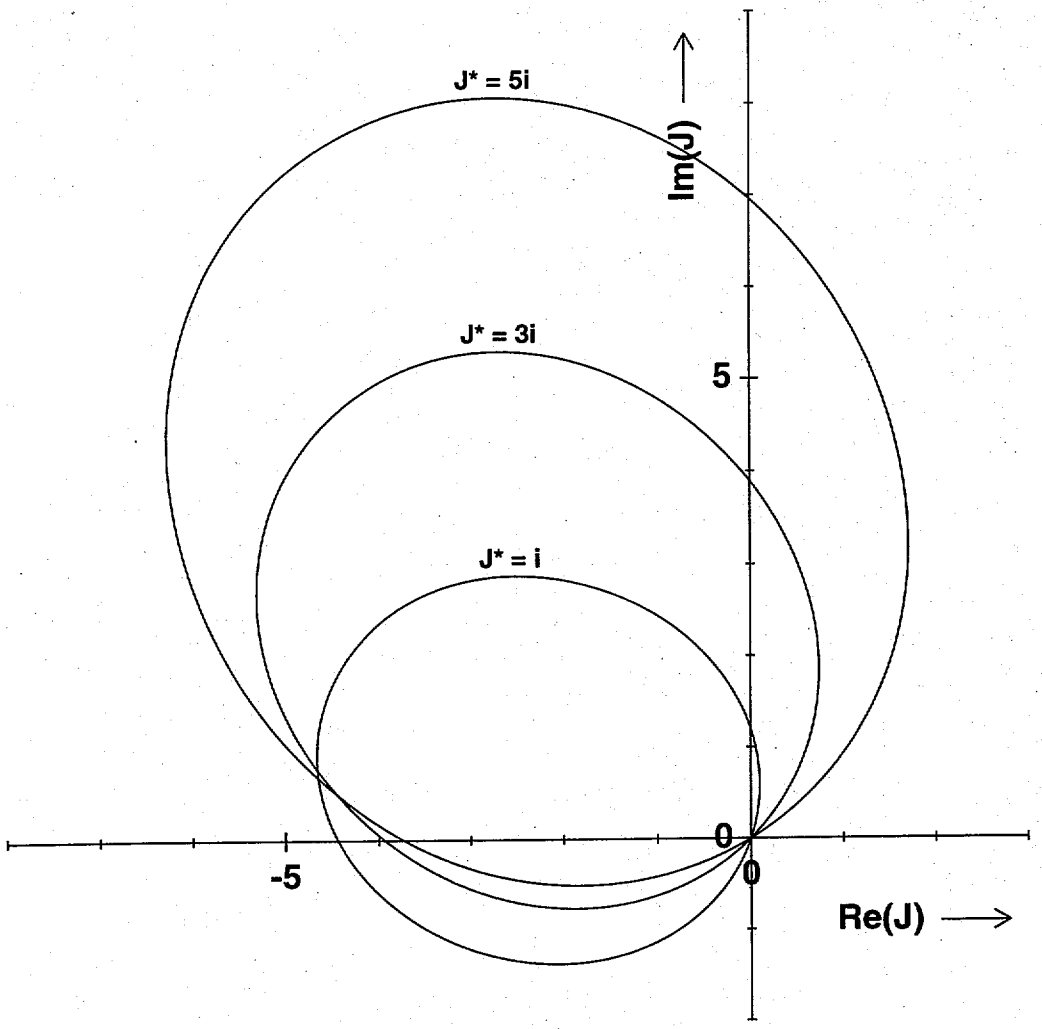

Figure 13. As in Fig. 10, but with $a_1 = a_3 = .2$ and $b = .5$

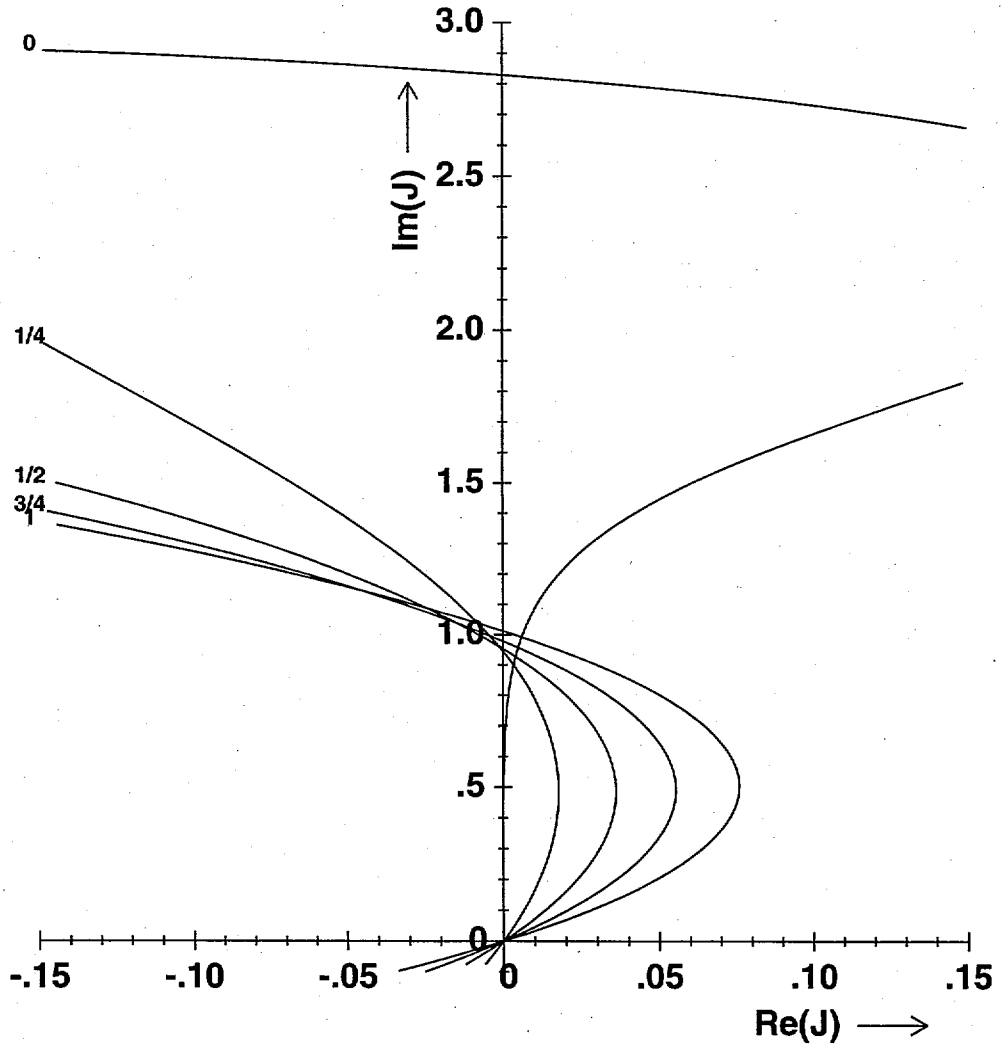

Figure 14. Stability boundaries for the modified Gill scheme with $a_1 = a_3 = 0.$ and $b = 1.$ but with the adjustment terms successively diluted by factors, $q = 1., .75, .5, .25, 0.$ The curve for the case, $q = 0,$ reverts to the stability curve for the explicit Runge-Kutta method. Note that the scale for the real axis is magnified more than the scale for the imaginary axis in order to better discriminate the stable and unstable regions of the imaginary J axis.

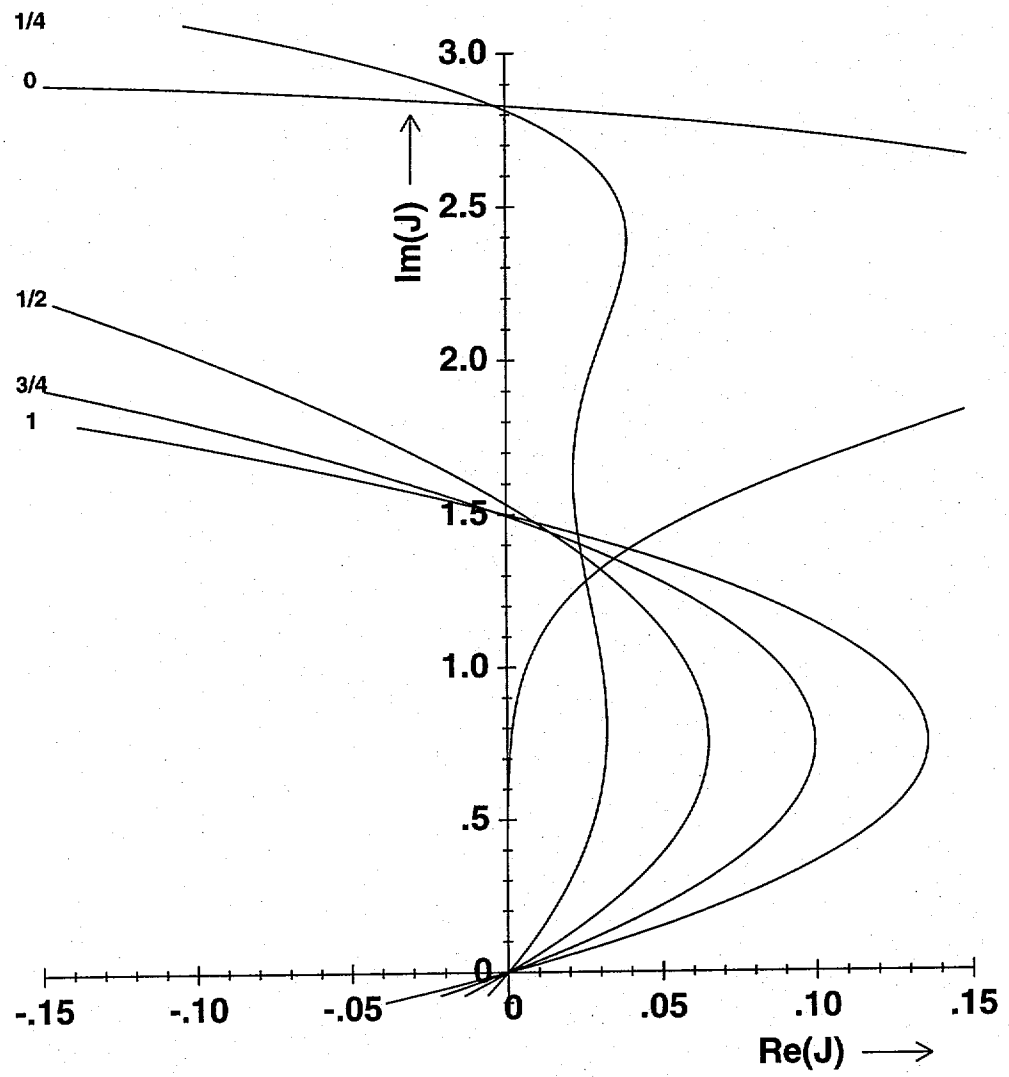

Figure 15. As in Fig. 14, but with $a_1 = a_3 = .5$ and $b = 0.$, i.e., the diluted form of the scheme shown in Fig. 12

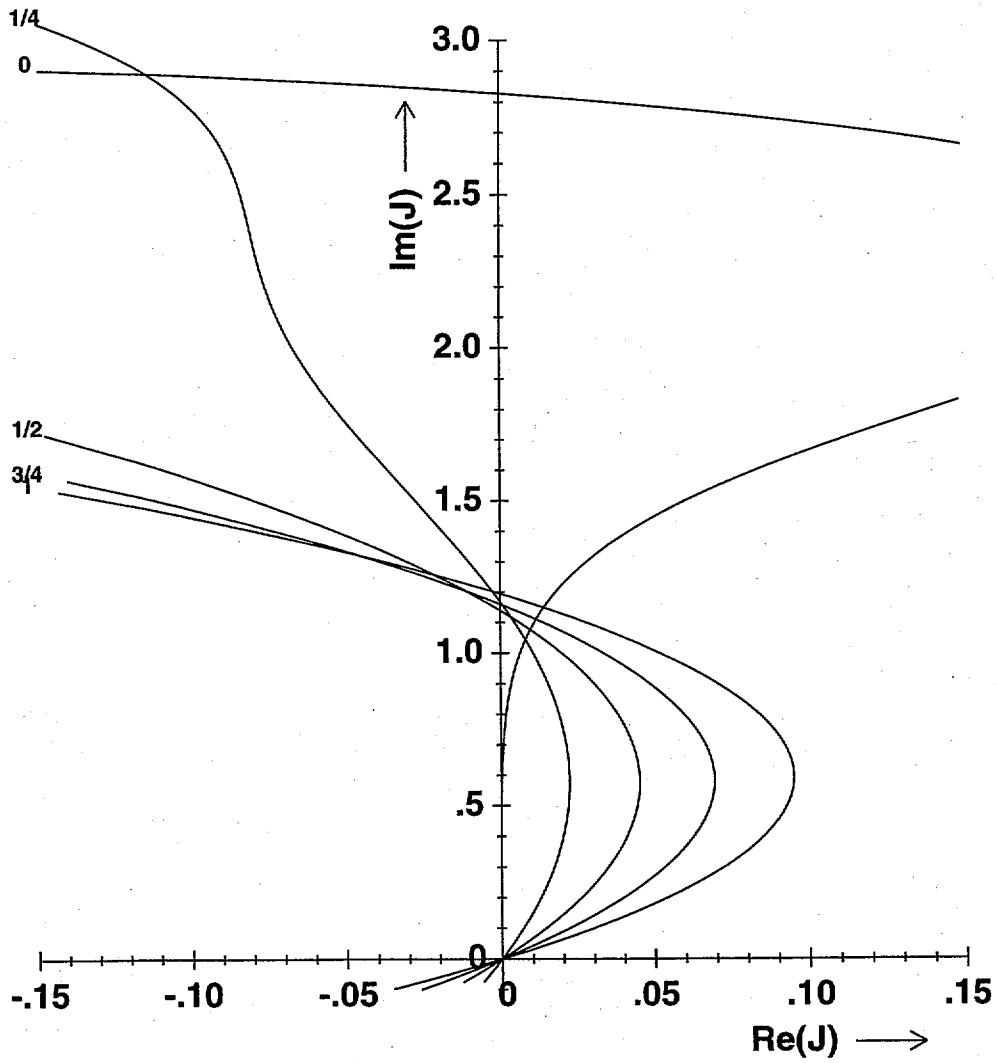

Figure 16. As in Fig. 14, but with $a_1 = a_3 = .2$ and $b = .5$, i.e., the diluted form of the scheme shown in Fig. 13

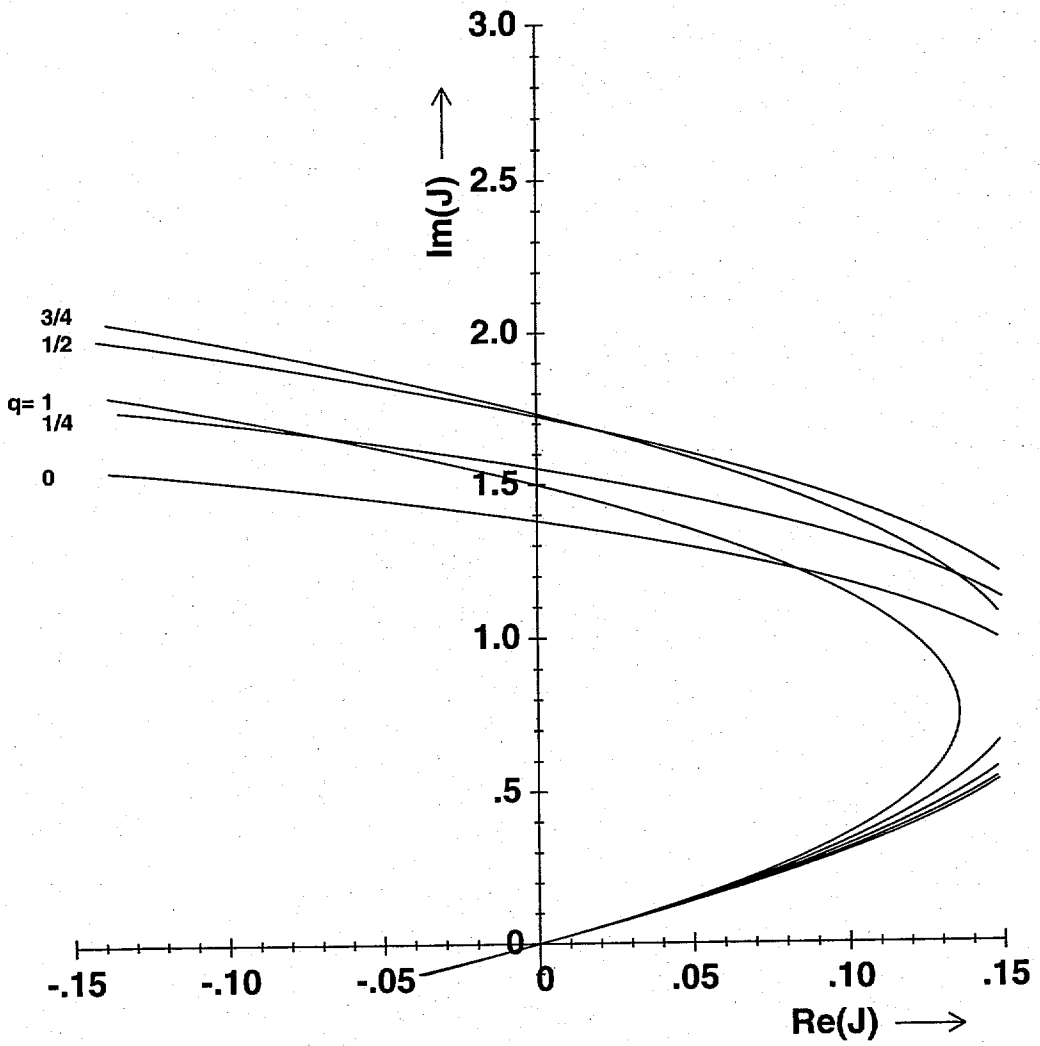

Figure 17. Stability curves for the modifies Gill scheme with $a_1 = a_3 = .5$ and $b = 0$, but with dilution factors q only applied to the adjustments of stages 2 and 4.

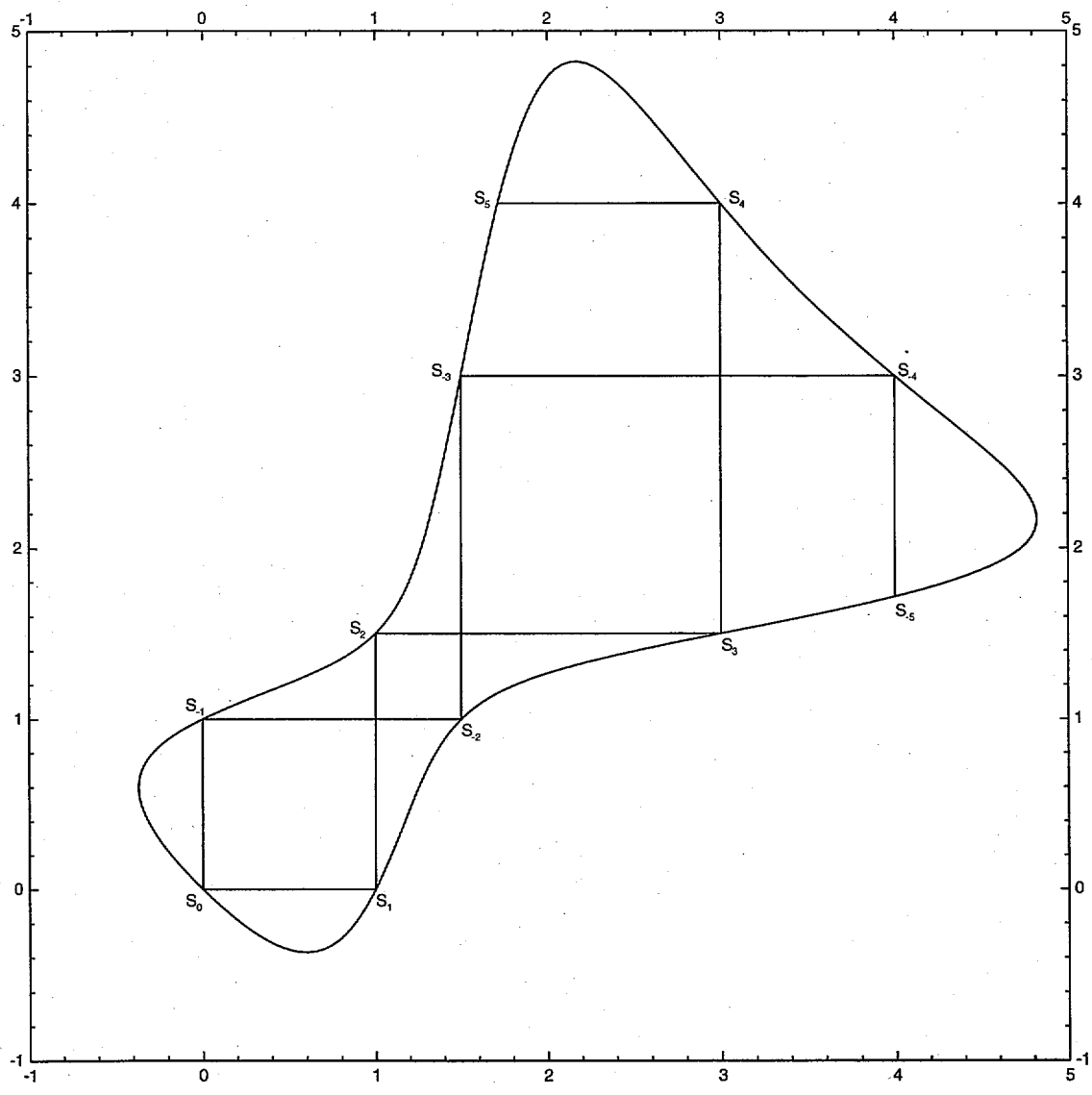

Figure 18. Locus of parameters, $X = 1/c_1$ and $Y = 1/(1 - c_2)$, of third-order Runge-Kutta schemes for which a low-storage implementation is possible. Some of the schemes with simple rational coefficients are included, together with a graphical depiction of the process which generates them as a double-sided sequence. The method recommended by Williamson corresponds to the scheme listed here as 'S_4'.